

University of Wollongong - Research Online

Thesis Collection

Title: Developing a subband model for blind signal separation in an acoustic environment

Author: Iain Trent Russell

Year: 2005

Repository DOI:

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Research Online is the open access repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

University of Wollongong Thesis Collections

University of Wollongong Thesis Collection

University of Wollongong

Year 2005

Developing a subband model for blind
signal separation in an acoustic
environment

Iain Trent Russell
University of Wollongong

Russell, Iain Trent, Developing a subband model for blind signal separation in an acoustic environment, PhD thesis, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, 2005. <http://ro.uow.edu.au/theses/553>

This paper is posted at Research Online.
<http://ro.uow.edu.au/theses/553>

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Developing A Subband Model for Blind Signal Separation in an Acoustic Environment

A thesis submitted in fulfilment of the
requirements for the award of the degree

Doctor of Philosophy

from

THE UNIVERSITY OF WOLLONGONG

by

Iain Trent Russell
Bachelor of Telecommunications Engineering (Honours Class I)

SCHOOL OF ELECTRICAL, COMPUTER
AND TELECOMMUNICATIONS ENGINEERING
2005

Statement of Originality

This is to certify that the work described in this thesis is entirely my own, except where due reference is made in the text.

No work in this thesis has been submitted for a degree to any other university or institution.

Signed

A handwritten signature in cursive script, appearing to read 'Iain Russell'.

Iain Russell

22nd August, 2005

Dedicated to my family

Acknowledgments

I would like to thank my supervisors, Dr. Jiangtao Xi, Prof. Joe Chicharo, and Prof. Alfred Mertins for their academic advice and continual support throughout the PhD. Gratitude is extended to Alfred and Jiangtao for making it an easy transition between supervisors over the course of the first two years of the project. I am also very grateful to Alfred for approaching me and providing the opportunity to do a PhD, providing the necessary scholarship and initial insight to the overall thesis topic.

Special thanks goes to Gaurav Srivastava and Dr. Mehran Abolhasan who as work colleagues always created an enjoyable atmosphere in the office and always fostered good discussions on topics ranging from women, politics, engineering, and religion.

Many thanks go to the members of TITR and the Whisper lab who made the last three years of hard work a pleasureable experience! The insight from different areas of research often provided a different way of approaching many problems specific to the area of BSS and was greatly appreciated. Special thanks also goes to Dr. Jason Lukasiak for his academic advice from time to time.

Finally I would like to express my deepest thanks to my parents, and family, for their encouragement, support and understanding over the last few years.

Author's Publications

Much of the work in this thesis has been published or has been submitted for publication as academic papers. These papers are:

1. Alfred Mertins and Iain Russell, "An extended ACDC algorithm for the blind estimation of convolutive mixing systems," in *Proceedings of Seventh International Symposium on Signal Processing and its Applications (ISSPA 2003)*, Paris, France, July 2003, vol. 2, pp. 527-530.
2. Iain Russell, Alfred Mertins, and Jiangtao Xi, "Time domain optimization techniques for blind separation of non-stationary convolutive mixed signals," in *Proceedings of 9th IASTED International Conference on Signal and Image Processing (SIP 2003)*, Honolulu, Hawaii, USA, August 2003, pp. 440-445.
3. Iain Russell, Jiangtao Xi, Alfred Mertins, and Joe Chicharo, "Blind separation of nonstationary convolutively mixed signals in the time domain," in *Proceedings of 7th International Symposium on DSP for Communication Systems (DSPCS03)*, Coolangatta, Qld, Australia, December 2003, pp. 93-98.
4. Iain Russell, Jiangtao Xi, Alfred Mertins, and Joe Chicharo, "Blind source separation of nonstationary convolutively mixed signals in the subband domain," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, Montreal, Canada, May 2004, pp. V-481-V-484.

5. Iain Russell, Jiangtao Xi, Alfred Mertins, and Joe Chicharo, "Integration of DFT and cosine-modulated filter banks with blind separation of convolutively mixed nonstationary sources," in *Proceedings of 3rd Sensor Array and Multichannel Signal Processing Workshop (SAM2004)*, Barcelona, Spain, July 2004, pp. CDRom.
6. Iain Russell, Jiangtao Xi, and Alfred Mertins, "Time Domain Blind Separation of Nonstationary convolutively mixed signals," in *Signal Processing for Telecommunications and Multimedia*, Vol. 27, Springer, New York, 2004, pp. 15-29.
7. Iain Russell, Jiangtao Xi, and Alfred Mertins, "Global optimization of uninitialized convolutive blind signal separation problems in the time domain," *Proceedings of 3rd Workshop on the Internet, Telecommunications, and Signal Processing (WITSP 2004)*, Adelaide, Australia, December 2004, pp. CDRom.
8. Iain Russell, Jiangtao Xi, Alfred Mertins, and Joe Chicharo, "Uninitialized sub-band blind signal separation of nonstationary convolutively mixed signals in acoustics using global optimization," Submitted to *IEEE Transactions on Speech and Audio Processing*.

Abstract

The focus of this thesis is to develop a framework for solving convolutively mixed blind signal separation problems in the subband domain. Current methods generally employ a discrete Fourier transform (DFT) to change the time domain convolutive model into many instantaneous multiplicative models to save on computations and convergence time. The motivation for approaching the problem from the subband domain is that there is an upper bound on the quality of separation for frequency domain methods where the mixing is done in a reverberant environment and there is a high number of unknown variables to solve for. This is shown with reference to the works in (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001; M. Ikram, and D. Morgan, 2000; R. Mukai, S. Araki, H. Sawada, and S. Makino, 2004). The model is developed throughout the thesis in a series of stages. Firstly we investigate modelling the convolutive Blind Signal Separation (BSS) problem completely in the time domain. The benefit of this is that by not performing any transforms we eliminate the local frequency permutation problem that is inherent in all convolutive BSS problems. To solve the permutation problem requires additional computational overhead. There is a tradeoff however according to how complex the mixing/demixing system is. The longer the reverberation time of an acoustic environment, the more unknown variables must be solved. The savings of performing multiplication in the frequency domain as opposed to convolution in the time domain must be compared to the savings of not doing the transform operator twice, as well as ensuring the local

permutation problem is solved.

Two new algorithms that avoid the local permutation problem are proposed and investigated. The first uses an alternating least squares approach (ALS) while the second uses joint diagonalization of output correlation matrices of the recovered signals. Where it is plausible to assume that we have some sort of a priori information that provides a good initial starting point for the unknown demixing system, then we only need to consider some type of local optimization procedure to solve the unknown demixing system. Two local optimization procedures investigated include the steepest gradient descent and Newton methods. Both types of local solvers were compared and the merits and disadvantages of each are specified in regards to the convolutive BSS time domain algorithm proposed. Where small convolutive mixing systems exist, such as in wireless communication mixing systems that assume a two ray model, the computational overhead that is increased by doing convolution in the time domain is offset more by the savings of not having to solve the local permutation problem and execute the transform operation.

In some cases, information pertaining to problem is unavailable. Geometric source separation assumes that there is some additional knowledge about the layout of the sensors with spatial reference to the source positions. This allows an angle of incidence of the sound wave impinging on the sensor array to either be known directly or calculated using various beamforming techniques. If we cannot assume to know such information, then multivariate complex problems with a high number of parameters become harder to solve for without getting spurious results from ill-convergence to local multiminima as opposed to the preferred global minima which corresponds to the desired demixing system that will allow signal separation. To avoid this, we integrate one of the proposed time domain convolutive BSS algorithms with a global optimization routine that is catered to suit the BSS convolutive problem model. A

branch and bound algorithm that uses division by hyper-rectangles is used to solve the uninitialized optimization BSS problem. With the validity of the proposed BSS time domain convolutive algorithm and the global optimization approach being justified, attention will then be focused on integrating these contributions into a model which uses subband decomposition before performing signal separation.

Various methods of subband decomposition are considered including using a uniform FIR analysis/synthesis filter bank based on DFT modulation as well as cosine modulation. The prototype window used is based on an extended lapped transform and was chosen due to the computational benefits of using lapped transforms. A framework for developing such a subband model is made with the main aspects of the model being the BSS algorithm and optimization approaches used, the way in which the observed signals from a multiple-input-multiple-output (MIMO) mixing system are decomposed via a filter bank, and the way in which the local permutation problem is overcome. In our work we propose a new subband detection, correction, and sorting routine for separated but arbitrarily permuted subbands over the entire spectrum.

Finally, a general and systematic approach for obtaining experimental measurements for generating the impulse response of an acoustic environment such as a typical office room, as well as the inverting MIMO system using wiener-hopf and optimal filtering theory is presented to allow full availability of information for the problem modelled in a practical environment as opposed to synthetic testing methods which are also examined.

Contents

1 Preliminaries	1
1.1 Introduction	1
1.2 Problem Statement	3
1.3 Blind Signal Separation: A Historical Review	8
1.4 Motivations and Applications	10
1.5 Approach and Contributions of this Thesis	12
1.5.1 Literature Review, Pre-Requisites and Outstanding Issues . .	12
1.5.2 Acoustic Modelling in Reverberant Environments	13
1.5.3 Time and Frequency Domain Convolutional BSS Algorithms .	13
1.5.4 Uninitialized BSS with Global Optimization	14
1.5.5 Subband BSS Model	15
1.6 Summary of Contributions in Order of Presentation	15
2 Literature Review, Pre-requisites, and Outstanding Issues	18
2.1 Introduction	18
2.2 General criteria for Signal Separability	19
2.3 Source Modelling	20
2.4 Mixing Systems	21
2.5 Measuring Performance	25

2.6	Stationary Instantaneous BSS	28
2.6.1	Higher Order Statistics	29
2.6.2	Estimation Theory	30
2.6.3	Information Theory	32
2.6.4	Global Scaling and Permutation Ambiguities	34
2.7	Stationary Convolutional BSS	35
2.7.1	Frequency Domain Methods	36
2.7.2	Local Scaling and Permutation Ambiguities	37
2.8	Non-stationary Sources	38
2.9	Subband Decomposition	39
2.9.1	Fundamental Concepts	39
2.9.2	Maximally Decimated Filter Banks and QMF	42
2.9.3	Block Transforms	45
2.9.4	Lapped Transforms	48
2.9.5	Extended Lapped Transform	50
2.9.6	Conclusion	53
2.10	Identifying Areas for Contribution	54
3	Acoustic Modelling	66
3.1	Introduction	66
3.2	Acquiring Room Response	67
3.3	Inverting the Room	71
3.4	Conclusions	76
4	Time and Frequency Domain Convolutional BSS models	77
4.1	Introduction	77

4.2	ALS Approach	79
4.2.1	Algorithmic Model	81
4.2.2	Simulation Results	86
4.3	Fullband TDBSS Approach	89
4.3.1	General Overview	89
4.3.2	Review of Instantaneous BSS	91
4.3.3	Proposed Convolutional TDBSS	93
4.3.4	Simulation Results	98
4.4	Conclusions	106
5	Uninitialized BSS with Global Optimization	107
5.1	Introduction	107
5.2	DiRect Algorithm	111
5.2.1	Convex Hulls	117
5.2.2	Clustering	117
5.2.3	Sequential Quadratic Programming	119
5.3	Simulation Results	121
5.4	Benefits for Small to Medium Scale Systems	124
5.5	Conclusions	126
6	Subband BSS Model	127
6.1	Introduction	127
6.2	Frequency Domain Methods and Limitations	130
6.3	Subband Model	133
6.3.1	Cosine Modulated FB	134
6.3.2	DFT Modulated FB	135

6.4	Integration of TDBSS into Subband Model	136
6.5	Comparing Subband and Frequency Domain BSS Models	138
6.6	Subband Coupling Metric	140
6.7	Dyadic Sorting Routine	142
6.8	Simulation Results	144
6.8.1	Benchmark ICA'99 Dataset	144
6.8.2	Synthetic Testing	145
6.8.3	Real Testing	151
6.9	Conclusions	153
7	Conclusions and Suggestions for Further Research	154
7.1	Conclusions	154
7.2	Suggestions for Future Research	157
A	Proof of Closed Form Analytical Expressions for Gradient and Hessian	169
A.1	Proof of Closed Form Analytical Expressions for Gradient and Hessian	169

List of Figures

1.1	General Mixing/Demixing system BSS model	4
1.2	\tilde{M} -channel uniform filter QMF bank (P. Vaidyanathan, 1993). . . .	5
1.3	BSS model of a two-input-two-output (TITO) system.	6
1.4	BSS model of a TITO system using subband decomposition.	7
2.1	Multipath Propagation in a room.	23
2.2	Multirate building blocks	40
2.3	Decimation by factor $R = 2$ in the time and frequency domain	40
2.4	Expanding by factor $R = 2$ in the time and frequency domain	41
2.5	Frequency responses of an \tilde{M} th band filter $H(z)$	44
2.6	Polyphase structure of \tilde{M} -channel QMF maximally decimated filter bank	45
2.7	Basis functions of block transforms and lapped transforms	48
2.8	Westner conference room photo (A. Westner, and V. Bove, 1999). . .	54
2.9	Westner conference room layout (A. Westner, and V. Bove, 1999). . .	57
3.1	A section of a typical MLS signal	69
3.2	Typical polar patterns for Shure SM57 (Shure Incorporated, 2004) . .	70
3.3	Digidesign Digi001 8 channel analogue I/O with 48kHz sampling. . .	70
3.4	Geometrical properties of room layout and direction of arrival of sources impinging on array manifold.	71

3.5	Measured room impulse responses with reverberation time of 200ms i.e. $P=1600$, for down-sampled rate of 8kHz for a TITO system. . .	72
3.6	System identification of the inverse FIR filter for a SISO system. . .	73
3.7	MIMO system identification for demixing FIR system using MMSE. . .	74
3.8	Wiener solution to TITO FIR demixing system with reverberation time of 250ms i.e. $Q=2000$, delay of for down-sampled rate of 8kHz. . .	75
3.9	Global TITO FIR system, i.e. cascaded mixing and demixing FIR systems.	76
4.1	Value of objective function on a logarithmic scale versus the number of full iterations.	87
4.2	Source power spectral densities (psd's) and their estimates. Legend: — psd1; - · - psd2; o estimate for psd1; x estimate for psd2. . .	88
4.3	Convergence of differing optimization methods for instantaneous BSS.	100
4.4	Convergence of gradient descent and Newton algorithms for a first order TITO FIR demixing system over 10 trials.	102
4.5	Convergence of Newton algorithms for first and third order TITO FIR demixing systems over 10 trials.	104
4.6	(a) and (b) are the two original signals, (c) and (d) are the convolatively mixed signals, (e) and (f) are the permuted separated results. . .	105
5.1	An iteration of the Direct algorithm for $n = 2$ dimensions. i.e. rectangles (D. R. Jones, C. D. Perttunen, and B. E. Stuckman, 1992) . .	116
5.2	An iteration of the Direct algorithm for $n = 3$ dimensions. i.e. rectangular prism (D. R. Jones, C. D. Perttunen, and B. E. Stuckman, 1992)	116
5.3	Hyper-rectangles on lower right convex hull (D. R. Jones, DIRECT, 2001)	119
5.4	Comparison of global and local optimization routines, glcCluster and snopt	123
5.5	(a) Above shows 'flops' using glcCluster and the method from (L. Parra and C. Spence, 2000) (b) Below shows is a magnification of (a). . .	125

6.1	General subband MIMO BSS model with oversampling factor $\frac{\tilde{M}}{R}$. . .	135
6.2	(a) Shows the impulse responses for the analysis filters for the first two subbands and (b) shows the magnitude frequency responses of the first three subbands for an ELT CM FIR FB when $\tilde{M} = 8$	136
6.3	Separation performance using three different BSS techniques for two TIMIT speech segments recorded with two cardioid microphones in a reverberant office environment.	140
6.4	Example of dyadic permutation sorting algorithm from (K. Rahbar, and J. Reilly, 2001) for separated subbands when total number of subbands $\tilde{M} = 8$	144
6.5	Virtual room synthetic mixing environment at 8kHz sampling frequency generated with the simroommix.m function.	146
6.6	Virtual room synthetic mixing impulse responses at 8kHz sampling frequency and reverberation time of 130 ms.	147
6.7	Wiener solution to TITO FIR synthetic demixing system with reverberation time of 200ms i.e. $Q=1600$, delay=32ms.	148
6.8	Ideal global TITO FIR system for virtual room.	150
6.9	First two separated adjacent subbands for $p = 1, 2$ when $\tilde{M} = 256, R = 128$	151

List of Tables

4.1	Closed form analytical expressions for the gradient and Hessian of the cost function and row-normilization constraint.	97
4.2	Gradient descent subband BSS algorithm for the joint-diagonalization task with a weighted constraint.	98
4.3	Newton-type subband BSS algorithm for the joint-diagonalization task with a weighted constraint.	98
5.1	Steps for algorithm "DIRECT" (M. Björkman and K. Holmström, 1999)	112
5.2	Parameters for algorithm "DIRECT" (M. Bjorkman and K. Holmström, 1999)	113
5.3	Steps for algorithm "DIRECT" (D. R. Jones, C. D. Perttunen, and B. E. Stuckman, 1992)	114
5.4	Steps for algorithm "Conhull" (M. Björkman and K. Holmström, 1999)	118
6.1	Separation performance comparison using SIR	151

List of Abbreviations

ABF	Adaptive Beamformers
AC	Alternating Columns
AIR	Acoustical Impulse Response
ALS	Alternating Least Squares
ANC	Adaptive Noise Cancellation
AR	Auto Regressive
ARMA	Auto Regressive Moving Average
BSS	Blind Signal Separation
CM	Cosine Modulated
CNV	Central Nervous System
DC	Diagonal Centers
DFT	Discrete Fourier Transform
DIRECT	Dividing Rectangles
DOA	Direction of Arrival
DSP	Digital Signal Processing
EEG	Electro-encephalographic
ELT	Extended Lapped Transform
EVA	Eigenvector solution for Blind Equalization
FB	Filter Bank
FIR	Finite Impulse response

FLOPS	Floating point Operations
GSM	Global System for MObile Communication
HOS	Higher Order Statistic
ICA	Independent Component Analysis
IID	Independent and Identically Distributed
IIR	Infinite Impulse Response
JADE	Joint Approximate Diagonalization of Eigenmatrices
KLD	Kullback-Liebler Divergence
LOT	Lapped Orthogonal Transform
LTl	Linear Time Invariant
MA	Moving Average
MAP	Maximum a Posteriori
MEG	Magneto-encephalographic
MI	Mutual Information
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MLS	Maximum Length Sequence
MLT	Modulated Lapped Transform
MRI	Magnetic Resonance Imaging
MSE	Minimum Squared Error
NP	Nonlinear Programming
NRR	Noise Reduction Ratio
PCA	Principle Component Analysis
PCI	Peripheral Component Interconnect
PI	Performance Index
PR	Perfect Reconstruction

QAM	Quadrature Amplitude Modulation
QMF	Quadrature Mirror Filter
QP	Quadratic Programming
QPSK	Quadrature Phase Shift Keying
RHS	Right Hand Side
SBSS	Subband Blind Signal Separation
SGD	Steepest Gradient Descent
SIR	Signal to Interference Ratio
SISO	Single Input Single Output
SNOPT	Sparse Nonlinear Optimization
SNR	Signal to Noise Ratio
SOS	Second Order Statistic
SQP	Sequential Quadratic Programming
SSARS	Source Separation Algorithm with Reference System
STFT	Short Time Fourier Transform
TD	Time Domain
TITO	Two Input Two Output

Chapter 1

Preliminaries

1.1 Introduction

Blind Signal Separation (BSS) is a field in Digital Signal Processing (DSP) which has developed over the last few decades into a broad and diverse method used to separate a linear or non-linear combination of mixed signals without any prior information of the mixing system, the source model or the number of sources and sensors. The field itself has been applied over the years in many different applications and although it has reached a level of maturity in the various algorithms used to implement specific models, the majority of cases result with a high computational complexity and long convergence times that impedes the growth of real-world applications that use BSS, for example real time applications. This is true especially in the area of convolutive BSS in reverberant environments where the signal of interest is speech or audio.

Research in convolutive BSS problems has lead to a variety of techniques. In most cases the algorithms used to model a solution to the convolutive BSS problem have been designed by exploiting the information in the observed mixtures in the frequency domain. There are limitations on the quality of separation with such methods and the need for additional assumptions on the problem to prevent ill-convergence to

a spurious local solution is required. The possibility of substantial gains in BSS is to merge the convolutive BSS algorithms that are robust, generic and computationally efficient with efficient subband techniques that effectively reduce the computational burden of solving one big problem by dividing it into smaller problems. At the same time it is also important to provide a framework to improve the quality of separation and eliminate or reduce the limitations that are inherent in the typical frequency domain approaches.

By investigating BSS in the subband domain, and extending algorithms to incorporate this aspect, a quantitative evaluation criterion can be developed to assess the validity of performing BSS in the subband domain as opposed to existing separation techniques.

This thesis considers several issues in relation to convolutive BSS problems in time, frequency, and subband domains. Firstly a general method for calculating the mixing responses of a multiple-input-multiple-output (MIMO) system in a reverberant environment is provided. In addition to the mixing responses, the corresponding inverse MIMO demixing responses as well as the corresponding subband components of the demixing responses is given. This allows full knowledge of the entire mixing/demixing system from a non-blind perspective and provides a benchmark of results to compare results derived from a blind approach to. Secondly, investigation of using different time and/or frequency domain algorithms to solve convolutive BSS problems whilst avoiding the local permutation problem is provided with the benefits of reduced computation for small to medium scale demixing systems, when using the proposed time domain algorithm(s), being provided. The next issue considered relates to optimization of the time domain algorithms. Specifically steepest gradient descent (SGD) and Newton methods are looked at for local optimization approaches where good initialization is assumed due to assumed prior knowledge of

the problem, e.g. spatial layout of sources with respect to sensors. In contrast, a global optimization approach is provided to allow separation where no initialization is required. Finally we merge all these approaches together into a subband based approach for large demixing systems that are exhibited in reverberant environments.

This chapter is organised as follows: Section 1.2 defines the proposed framework for subband BSS when dealing with convolutive mixing systems in reverberant environments. Section 1.3 gives a brief historical review of BSS techniques. Section 1.4 provides the motivations for the proposed research. The approach and contributions of this thesis are described in Section 1.5. Finally, Section 1.6 lists the contributions in point form.

1.2 Problem Statement

To understand the proposed framework for subband BSS, it is necessary to understand the general concepts of BSS and multi-rate filter banks. Each field will be briefly defined individually as part of the problem statement and then will be integrated to a more general problem statement for performing BSS in the subband domain. A more thorough analysis of varying BSS algorithms and concepts relevant to the work proposed in this thesis is provided in Chapter 2.

The problem of BSS is a very general and fundamental one. Related fields using similar techniques include Independent Component Analysis (ICA), which has also received a great deal of attention over a similar time span, as well as Principle Component Analysis (PCA), deconvolution, and blind equalization to name just a few. All of these related fields are specific instances that could be derived from the more generalized BSS problem. Figure 1.1 illustrates a generic description of the BSS problem.

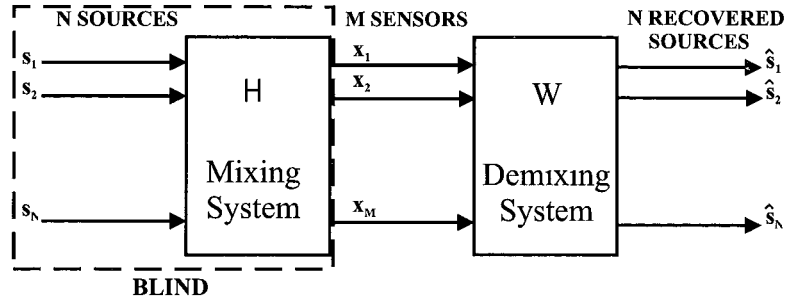


Figure 1.1 General Mixing/Demixing system BSS model

Suppose there are M observed signals x_1, x_2, \dots, x_M that are linear or non-linear combinations of N independent source signals s_1, s_2, \dots, s_N . The aim of BSS is to produce N outputs $\hat{s}_1, \hat{s}_2, \dots, \hat{s}_N$ that recreate the original source signals up to some arbitrary delay. The only assumption that can be made about the sources is that they are statistically independent which is the fundamental criterion behind BSS and will be defined in Chapter 2.

The term *blind* implies that the recovered signals are determined without any prior knowledge of the source signals or the combination system that mixes them. The two basic types of sources can be classified as independent and identically distributed *i.i.d* sources i.e. stationary sources, and non-stationary sources. The N source signals are mixed or filtered by some arbitrary unknown system H . H is usually represented by an $M \times N$ matrix of either scalars for the instantaneous case, or finite impulse response (FIR) filters in the case of convolutive mixing. The observed or mixed signals may or may not have additional noise added to them depending on the complexity of the model. In most real-world applications the addition of noise is usually the case but we consider the noiseless case for simplicity.

BSS basically calculates the demixing system, or $N \times M$ matrix W , so that when mixed with the observed signals, recovered signals are obtained that match the un-

known source signals up to an arbitrary global permutation and scaling factor.

An understanding of filter bank theory, multi-rate systems and lapped transforms is essential to deciding on what kind of subband decomposition provides the most optimal performance in the context of BSS. A more detailed analysis of the designs and fundamental building blocks of efficient multi-rate filter banks and block transforms will be provided in Chapter 2 but a general formulation of an \tilde{M} -channel multi-rate uniform filter bank system is illustrated with the aid of Figure 1.2.

The \tilde{M} -channel uniform filter bank is an efficient way of attempting to study the general theory of alias cancellation and perfect reconstruction (PR), two vitally important concepts in design criteria for multi-rate systems. A polyphase implementation of the filter bank, shown above in direct form, would be generally used for implementation due to computational efficiency however, for the purposes of defining the problem statement of the thesis the simplest approach is taken.

In Figure 1.2 the signal $x(n)$ is split into \tilde{M} subband signals by the \tilde{M} analysis filters $H_p(z)$. Every signal is then decimated by a sub-sampling factor R to obtain the subband signals. These subband signals may be subject to some sort of subband processing as commonly found in subband coding for images and speech. The processed

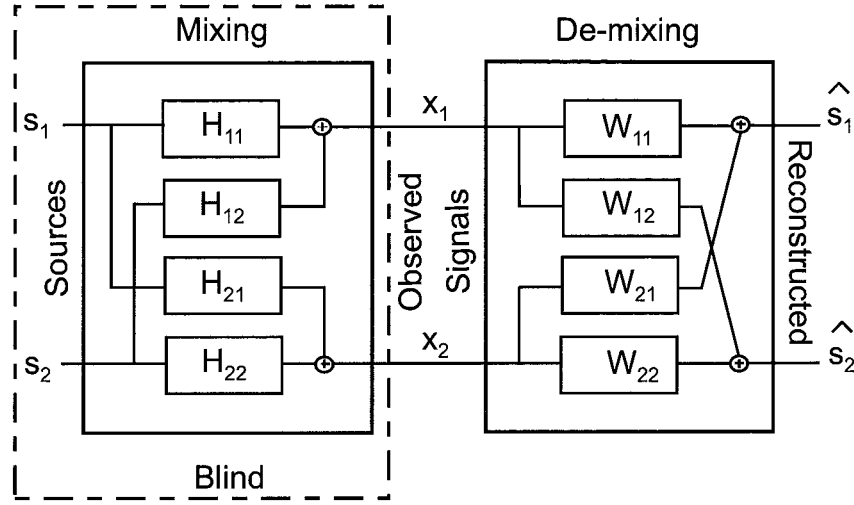


Figure 1.3 BSS model of a two-input-two-output (TITO) system.

signals are then passed through R -fold expanders before being recombined via the synthesis filters $F_p(z)$ to produce the reconstructed signal $\hat{x}(n)$ (P. Vaidyanathan, 1993).

The reconstructed signal $\hat{x}(n)$ will not be identical to $x(n)$ due to errors created by the filter bank system such as aliasing, imaging, amplitude distortion, and phase distortion which will be discussed in Chapter 2. To overcome and minimise these errors requires careful design of the analysis and synthesis filter banks. PR-QMF and pseudo QMF FIR filter banks using cosine modulation have been developed independently by Koilpillai and Vaidyanathan (Koilpillai and Vaidyanathan, 1992) and Malvar (H. Malvar, 1990) in the early nineties and have several advantages, which will be discussed in Chapter 2. The theory presented on these designs also forms the basis for using Lapped Transforms in the design of the prototype function from which all impulse responses of analysis and synthesis filters can be derived. This will be discussed in Chapter 2.

With both areas of BSS and filter banks briefly described, a description of how these

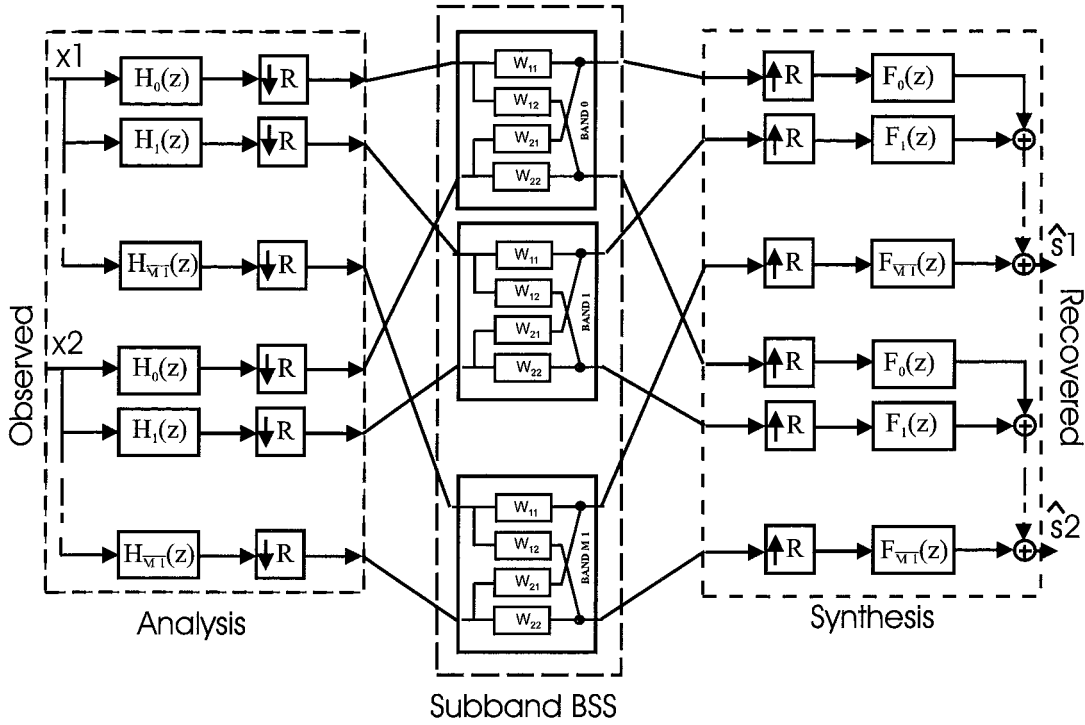


Figure 1.4 BSS model of a TITO system using subband decomposition.

subjects can be integrated to form the basis of research on the framework for BSS in the sub-band domain can now be addressed. The problem statement can be described with the aids of Figures 1.3 and 1.4.

Figure 1.3 shows normal BSS implementation for a TITO system. This describes a system where we assume there are two sources and two sensors. The majority of BSS algorithms developed in current literature typically use a TITO system, whether it uses synthetic or real data simulations, to conduct a performance and comparison evaluation with existing BSS algorithms. Obviously a MIMO system describes the general case where the number of sources and sensors is unknown and most algorithms are extended from the TITO case to the MIMO case theoretically. Despite this, actual implementations of these algorithms for the MIMO case are not investigated exhaustively and so for comparison the TITO case is used. Figure 1.3 basically

represents what could be referred to as direct BSS in the fullband domain.

Figure 1.4 shows the general approach to extending BSS to the sub-band domain, which is fundamental to this research initiative. The signals coming from different sensors are first decomposed via filter banks into narrow-band signals, and then the separation is carried out for narrow-band signals. Finally, the individual solutions are synthesized to yield the overall solution. As can be seen, the determination of the demixing system \mathbf{W} in the subband domain as opposed to a fullband domain will save on the number of computations as the demixing system increases in dimension. Things investigated in this thesis include designing effective time domain BSS algorithms to solve BSS problems without transforming to the frequency domain and thus eliminating the inherent local frequency permutation problem. An analysis of which filter bank design works best and is most efficient is given. Whether oversampling, or critical sampling should be used, uninitialized global optimization techniques for convergence to global minimums and any possible problems that may arise including aliasing, signal recovery and the permutation problem that arises in BSS are also investigated.

1.3 Blind Signal Separation: A Historical Review

The seminal work on source separation was given in a meeting for Neural Networks for Computing in Utah, 1986. Jeanny Herault and Christian Jutten contributed a research paper (J. Herault, and C. Jutten, 1986) that presented a recurrent neural network model and a learning rule that could blindly separate mixed combinations of independent signals. The algorithm stemmed from how the central nervous system is able to differentiate independent signals from mixed signals with the independent signals, describing some arbitrary body process or movement. The only assumption

that was made was source independence. This approach was explained and further developed by Jutten and Herault, and Comon (1991), Cichocki and Moszczynski (1992), and others. In 1994 Comon contributed a more generic examination of source separation and Independent Component Analysis (ICA) that crystallized Herault and Jutten's findings. He looked at cost functions that minimized the mutual information between the sensor signals. In the signal processing community an abundance of algorithms were being formulated based on cumulants during the early nineties when research on neural networks was also very popular.

Bell and Sejnowski (A. Bell, and T. Sejnowski, 1995) were the first researchers to investigate and explain the blind source separation problem from an information-theoretic perspective. They used adaptive methods that proved more plausible from a neural processing perspective than the cumulant based cost functions proposed by Comon. During the same time, similar algorithms based on different approaches were proposed. Gaeta and Lacoume (1990) introduced the maximum likelihood estimation approach; Harhunen and Joutsensalo (1994) developed nonlinear PCA, Girolami and Fyfe proposed negentropy maximization (1996), while Lee and Sejnowski (1997) provided a unifying framework to the BSS problem by describing the relationships between the different algorithms. Most algorithms involve defining some objective function to minimize or maximize, where the solution corresponds to signal separation.

A lot of the initial investigation done on BSS utilised knowledge of ICA where the mixing environment was an instantaneous one and in most cases only a two-input-two-output (TITO) or single-input-single-output (SISO) system. Yellin and Weinstein addressed the multichannel BSS problem (D. Yellin, and E. Weinstein, 1994). To address the problem of convolutive mixtures for real environments a theoretical framework was described by Lambert (R. Lambert, and A. Bell, 1997). The mul-

tipath extension of blind source separation methods can be seen in the frequency domain using FIR matrix algebra. More recently the approach of using Second Order Statistics (SOS) solely as opposed to a combination of Higher Order Statistics (HOS) and SOS has been investigated. The joint diagonalization approach of multiple cross-correlation matrices over multiple time instances involves exploiting non-stationary sources such as speech for example, which is quasi-stationary over a short period of time. Recent research on solving BSS problems in a reverberant convolutive mixing environment using frequency domain approaches, including papers by (L. Parra and C. Alvino, 2002; L. Parra and C. Spence, 2000; K. Rahbar, J. Reilly and J. Manton, 2004; K. Rahbar, and J. Reilly, 2001; H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004a), and (H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004b), gives new ideas to solving convolutive BSS problems, however such methods have additional assumptions and certain limitations which will be described in Chapter 2.

These are just a few of the important contributions made to the field of BSS in the last decade. A more detailed examination of some of the findings that have been made in the broader areas of instantaneous and convolutive mixing will be given in Chapter 2 along with a more detailed analysis of current BSS algorithms that assume non-stationary signals and the use of SOS to achieve separation.

1.4 Motivations and Applications

Over the last ten years the problem of BSS has been investigated extensively with motivations generated by a wide variety of interests and applications including, higher order statistics, neural networks and artificial learning, noise cancellation, array beam forming, speech and image enhancement and recognition, communications over unknown channels including fading in mobile communications, biomedicine and neu-

rology.

Herault and Jutten conducted the originating work to the BSS problem (J. Herault, and C. Jutten, 1986). The motivation for the research came from the central nervous system's ability to separate signals. They proposed an adaptive algorithm in a simple feedback architecture found at various levels of the central nervous system. The learning rule was based on a neuromimetic approach and was able to simultaneously separate unknown independent sources using multiple sensors. This idea has since been extended in the biomedical field to the capturing of electro-encephalographic (EEG) and magneto-encephalographic (MEG) data. Electrode sensors placed noninvasively on the surface of the head or body obtains a mixed combination of process signals. A linear decomposition of the observed data via BSS methods can recover the electrophysiological data for analysis of brain or neuronal processes (S. Makeig, A. Bell, T. Jung, and T. Sejnowski, 1996).

Adaptive Noise Cancellation (ANC) another application of BSS is a particular case of the TITO system i.e. 2 sources, 2 sensors separation system. The applications of ANC include the removal of engine noise from the cockpit of an aircraft, the removal of cross talk between adjacent communication channels as well as speech enhancement.

Some of the most commercially viable applications in BSS, which motivates a great deal of research in the area, are from the fields of speech recognition and enhancement and human acoustics. Practical systems such as hearing aids, speech recognisers and teleconferencing facilities are possible real world applications for BSS. One of the main attractions to research on BSS lies in attempting to solve the Cocktail Party Problem. In this problem there are several competing speakers or sources of speech signals. The difficulties in solving the cocktail party problem have been in-

investigated by Torkkola in (K. Torkkola, 1999). One of the main difficulties is the fact that in real acoustic environments such as rooms they are highly reverberant requiring very long FIR demixing filters in the order of thousands of taps. The algorithms developed for signal separation are too computationally intensive to deal with such situations. If a different approach can be taken to reduce the complexity of models based on real variables then such problems could be solved in practice more efficiently.

Numerous BSS algorithms have been demonstrated to be effective in a variety of situations and the theoretical derivations of models are well understood. However, implementation of many solutions is highly inefficient and computationally expensive. By incorporating a sub-band approach this research hopes to improve upon the existing separation performance of typically used frequency domain approaches.

1.5 Approach and Contributions of this Thesis

This thesis attempts to improve the quality of separation for convolutive BSS problems where mixing systems range from a small number of dimensions to a relatively large number in the case of highly reverberant environments.

A general framework and methodology for deriving a subband based BSS model is proposed as the theme of the thesis. Each central idea in the thesis serves to contribute to the overall subband model and improve separation performance over a typical frequency domain method.

1.5.1 Literature Review, Pre-Requisites and Outstanding Issues

A comprehensive literature review provides current research in the area and identifies potential areas for contribution. The background provides the foundation knowl-

edge of terms and concepts that are necessary to understand before examining the proposed contributions including algorithms, methods, and frameworks that are presented in this thesis.

1.5.2 Acoustic Modelling in Reverberant Environments

When dealing with BSS of convolutively mixed nonstationary sources such as speech in a reverberant environment, a superior comparative performance is achieved if all information about the BSS problem is readily available or easily produced through a series of steps.

In most cases, current literature does not make impulse responses of mixing channels available but instead provides geometrical layouts of rooms used to conduct experiments or provides observed recordings. If the experiment is treated as a non-blind problem first, adequate information can be obtained to see how well the blind algorithm performs.

Chapter 3 proposes a method to obtain all relevant information for a MIMO BSS mixing/demixing system in a reverberant environment where convolutive mixing takes place in the form of multipath propagation. Firstly a method to obtain the MIMO mixing system is defined as well as the process to find the corresponding demixing impulse responses. It serves to provide a methodology for conducting experiments that investigate BSS of convolutively mixed non-stationary speech/audio signals in a reverberant environment.

1.5.3 Time and Frequency Domain Convolutive BSS Algorithms

The majority of problems that consider convolutive BSS assume a transformation to the frequency domain. Although this reduces computational complexity and parallels motivation for a subband BSS model, there are certain limitations on separation

performance of frequency domain methods when impulse responses have thousands of coefficients due to rooms with long reverberation times. These will be discussed in Chapter 2. For both the typical frequency domain method and proposed subband method, the local frequency permutation and scaling problems exist and must be rectified before either taking the inverse fourier transform, or synthesizing in the filter bank. However if the length of the impulse response is short enough, for example communication channels that use a two-ray model, then there is incentive to formulate some BSS criteria solely in the time domain which eliminates the overhead required for solving the local frequency/subband permutation problem.

Chapter 4 provides a new time domain BSS algorithm that solves convolutive BSS problems that assume non-stationary sources, completely in the time domain. It also investigates a time-frequency domain alternating least-squares (ALS) BSS algorithm that overcomes the local permutation problem. Mitigating the local permutation problem effectively reduces computational overhead for smaller to medium sized systems. A pure time domain implementation of BSS for convolutive mixing is only beneficial up to a certain number of dimensions of the unknown system when compared to a typical frequency domain method. Both proposed algorithms employ joint diagonalization and SOS and assume non-stationary input signals. One method uses an alternating least squares (ALS) approach for optimization while the other uses Newton and gradient descent methods initially, and then is further extended to use a global optimization approach.

1.5.4 Uninitialized BSS with Global Optimization

In Chapter 5 we propose a global method of optimization as opposed to typically used local optimization methods for solving multivariate non-linearly constrained problems. The main benefit of using a global optimization method is that there is no

requirement for a good initial starting value for the unknown demixing system. This is necessary when using local optimization algorithms when the function is multi-modal to prevent ill-convergence to incorrect local minima. This chapter highlights some of the different global optimization methods available and customizes a branch-and-bound algorithm to solve the convolutive BSS problem in the time domain for one of the proposed time domain algorithms. A comparison is made to an initialized local frequency domain method for small to medium scale systems.

1.5.5 Subband BSS Model

Chapter 6 extends the proposed time domain BSS algorithm into the subband domain. An understanding of the theory of filter-banks, multirate systems, block transforms and lapped transforms is required and is given in Chapter 2. Chapter 2, in addition to providing a literature review, aims to address the fundamental building blocks and concepts that need to be addressed when designing efficient filter-bank systems, especially in the context of BSS. Chapter 6 integrates the global optimization, cosine modulated filter-bank model, and one of the proposed time domain BSS algorithms from Chapter 4. This chapter culminates all concepts discussed throughout this thesis prior to Chapter 6 and provides a general framework of modelling BSS problems in the subband domain. Also a new method for solving the subband permutation problem is given as well as a dyadic sorting routine for all separated subbands.

1.6 Summary of Contributions in Order of Presentation

- A methodology for obtaining all relevant information in a non-blind sense for a MIMO mixing/demixing system in a reverberant environment to serve as a

benchmark for comparing blind results to for quality analysis.

- An extension to the ACDC (ALS) algorithm introduced by (A. Yeredor, 2002) is made from the instantaneous case to the convolutive case with non-white sources.
- An extension of the joint-diagonalization algorithm for the instantaneous mixing case introduced by (M. Joho and K. Rahbar, 2002) is made to the convolutive case in the time domain.
- Closed form analytical expressions for gradient and Hessian are derived for the time domain convolutive BSS algorithm proposed for both the objective and constraint functions.
- The benefits of avoiding the local permutation problem and solving convolutive BSS problems completely in the time domain when the demixing system has a small to medium number of dimensions.
- Using global optimization to solve the constrained objective criterion of the time domain BSS algorithm proposed.
- Evaluating the benefits of not having to assume *a priori* information to obtain good initial values for the unknown demixing system when using a global optimization method as opposed to currently used local methods.
- Integrating the existing proposed time domain algorithm with a subband decomposition of the observed signals and completing the separation phase in the subband domain for each relevant subband. This framework is proposed for convolutive mixing BSS problems where the number of dimensions for the unknown demixing system is relatively high due to long reverberation times in the mixing environment.

- A new subband permutation mechanism for detecting and correcting separated subbands that have differing permutations. Also combined with this is a dyadic tree sorting routine.

Chapter 2

Literature Review, Pre-requisites, and Outstanding Issues

2.1 Introduction

For blind signal separation (BSS) of non-stationary convolutively mixed signals in a reverberant environment, an understanding of some of the fundamental concepts for BSS theory and subband decomposition is necessary. This chapter firstly defines the underlying assumption that the majority of BSS algorithms make on the sources and that is of statistical independence. The various ways to model the unknown sources is briefly discussed focusing on the two main categories of stationary and non-stationary sources. The type of algorithmic model used to solve various BSS problems, will depend on how the sources are modelled, which will depend on the targeted application. Both types of general mixing models will also be defined starting with the simple instantaneous model and then extending it to the more realistic convolutive mixing model. A brief discussion on some of the main approaches to instantaneous and convolutive BSS problems will be given with particular emphasis on convolutive methods that exploit second order statistics (SOS) via joint diagonalization with non-stationary sources. A general description of how convolutive

BSS problems are approached when using a frequency domain method will be also given highlighting some of the fundamental problems including the *local permutation* problem and limitations on separation performance. Various quantitative ways to measure the quality of separation will be also defined. A review of filter bank theory, lapped transforms, and cosine modulation is provided and a discussion is given on the adverse-effects that blind signal separation introduces in the subband domain and how to overcome them.

Finally, a thorough investigation on current literature in the area of convolutively mixed BSS problems that assume non-stationary sources is given with particular emphasis on audio/acoutical applications of the problem. The purpose of this final section is to provide a summary of what current research is being done in the area of blind signal separation for convolutive mixing in time, frequency, and subband domains for nonstationary input signals such as speech. Outstanding issues which have not been addressed in the areas examined are identified and provide a valid justification for the contributing work proposed in this thesis.

2.2 General criteria for Signal Separability

A critical assumption made for all BSS algorithms is that of the sources having statistical independence. Even if nothing is known about the distributions or the parametric family that the sources belong to, information concerning the joint distribution of the signals can be provided by the mutual independence assumption. Although this assumption seems generic and simple, it is this very attribute, which provides a robust theoretical framework for BSS in all its variant forms; the weaker the assumption the wider the applicability.

A set of random variables y_i for $i \in 1, 2, \dots, m$ are defined to be independent if

their joint distribution is a product of their marginal distributions. To further define this concept of independence, consider the case of two random variables y_1 and y_2 . The variables y_1 and y_2 are considered independent if information on the value of y_1 doesn't provide any information on the value of y_2 , and vice versa. This is the case for the sources but not the mixed signals. The joint probability distribution function $p(y)$ of the signals can be expressed as,

$$p(y) = p_1(s_1) \times p_2(s_2) \times \dots \times p_n(s_n) = \prod_{i=1}^n p(y_i) \quad (2.1)$$

BSS techniques separate the observed signals through establishing statistical independence between the output signals. Various techniques to separate the mixed signals into independent recovered versions of the assumed independent source signals include maximizing the uncorrelatedness or nongaussianity of the mixed signals after applying an appropriate separation transform, kurtosis minimization or maximization, negentropy, mutual information, maximum likelihood, infomax and other information-theoretic approaches, and joint-diagonalization approaches. Usually these fundamental approaches derive some objective function that may be subject to constraints. This objective usually is optimized through a local or global optimization method. Additional criterion, depending on the information that is known *a priori*, can be made depending on the type of application or environment that BSS will be performed in.

2.3 Source Modelling

In BSS there are basically two general categories that sources can be divided into, stationary and non-stationary. Most BSS algorithms that exploit not only second order statistics (SOS) but also higher order statistics (HOS) to achieve separation usually assume stationary source signals. For stationary sources additional constraints are

needed requiring more than SOS. SOS allows decorrelation but HOS allows separation when source signals are stationary. As the primary interest is in separating mixtures of speech in a reverberant environment, the research covered in this thesis focuses more solely on source signals that are assumed to be non-stationary such as speech which is considered quasi-stationary over a period of 20 ms. A brief review of BSS approaches that employ HOS of the observed signals is provided, however more attention will be reserved for BSS methods that use joint-diagonalization.

Non-stationary sources have distributions that vary in the temporal domain. Non-stationarity may arise either from the source signals themselves (such as speech), or from channel impairments (such as fading in wireless communications channels) (B. Krongold and D. Jones, 2000). Separation algorithms that assume non-stationary source signals do not need to use the computationally demanding HOS required by i.i.d source signals. A set of SOS does not usually provide enough constraints to allow separation for stationary signals, however with non-stationary signals the SOS of the signals are changing with time and can provide enough information to allow separation using differential correlation. A particular non-stationary BSS algorithm that separates speech that will be examined for comparison with the proposed algorithms in the time and subband domains is (L. Parra and C. Spence, 2000). This algorithm will be reviewed in further detail in Chapter 4 but basically exploits the time-varying nature of speech.

2.4 Mixing Systems

The simplest BSS mixing model assumes that there are N statistically independent source signals $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ which are combined in a linear and instantaneous fashion to produce the observation of M mixtures $\mathbf{x}(t) =$

$[x_1(t), x_2(t), \dots, x_M(t)]^T$, where $M \geq N$. This combination is described mathematically as,

$$\mathbf{x}_i(t) = \sum_{j=1}^N h_{ij} s_j(t) \quad \text{for } i = 1, 2, \dots, M. \quad (2.2)$$

In this model we assume that the unknown source signals $s_j(t)$ are multiplied by scalar values h_{ij} and added to produce the mixed signals $x_i(t)$. In Figure 1.3 the unknown impulse responses that represent the coupling of source j to sensor i are given as FIR filters. However for the case of linear instantaneous mixing these are simply scalars or zero order FIR filters of length one. This makes the mixing process considerably simpler. This model can also be interpreted more compactly in vector matrix form as,

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t), \quad (2.3)$$

where,

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1N} \\ h_{21} & h_{22} & \dots & h_{2N} \\ & & \dots & \cdot \\ h_{M1} & h_{M2} & \dots & h_{MN} \end{bmatrix}, \quad (2.4)$$

and $\mathbf{x}(t)$ and $\mathbf{s}(t)$ are the column vectors as previously defined. The instantaneous BSS problem exists in recovering the source vector $\mathbf{s}(t)$ using only the observed data $\mathbf{x}(t)$. The BSS problem for the instantaneous case can be formulated as the computation of an $N \times M$ matrix \mathbf{W} whose output,

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t), \quad (2.5)$$

is an estimate of the source signal column vector $\mathbf{s}(t)$. There are many different BSS models for determining the separating matrix \mathbf{W} that should ideally be the inverse of the scalar mixing matrix \mathbf{H} . The simplicity of assuming instantaneous mixing is that BSS algorithms can be developed easily however in many cases the instantaneous mixing of signals is only a synthetic phenomenon. Replication of mixing

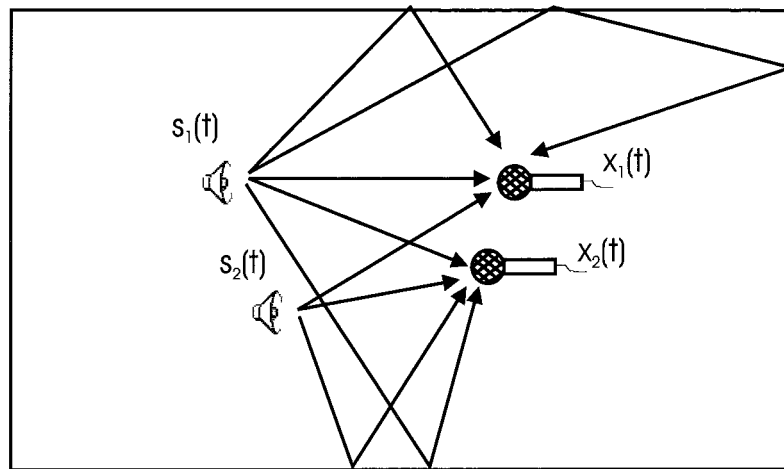


Figure 2.1 Multipath Propagation in a room.

conditions of systems in the real world is more precisely described by convolutive mixing systems such as FIR filters. Consider the multipath propagation of sound in a reverberant acoustic environment such as a typical office room. Multiple copies of each source signal will arrive at the sensor(s) at delayed intervals. If BSS models assuming instantaneous mixing conditions were used this would only take into account the first copy that arrives at the destination sensor, i.e. the direct path, and would not provide a true model representative of the real world application.

Although an understanding of how BSS algorithms are developed in an instantaneous climate is important, algorithms that model realistic scenarios need to be extended to separate convolutively mixed source signals. A convolutive combination system accounts for the properties of the multipath transmission of signals in a real reverberant environment. Figure 2.1 illustrates a simple example of multipath propagation in a room for convolutive mixing.

Multipath propagation introduces different propagation delays to the destination microphones and signal distortion from signal reflection off physical objects with different damping properties. Each source signal will experience filtering between itself

and the respective sensors. The filtering properties can be described mathematically by FIR filters with length P ,

$$\mathbf{x}(t) = \sum_{\tau=0}^{P-1} \mathbf{H}(\tau) \mathbf{s}(t - \tau). \quad (2.6)$$

If there are N sources and M sensors or microphones then there will be $N \times M$ different impulse responses representing each of the different propagation or transmission channels. Equation (2.6) can also be written more compactly as,

$$\mathbf{x}(t) = \mathbf{H}(t) * \mathbf{s}(t). \quad (2.7)$$

The convolution operator $*$ is used above. The best type of filter to model the impulse responses of the direct and cross-channels is the FIR filter. FIR filters are Moving Average (MA) systems comprised of a polynomial in the z domain. Alternatively there are Auto Regressive (AR) systems that are all poles or ARMA systems that are a combination of both poles and zeros. BSS algorithms rely on finding a way of finding the inverse of the mixing system to perform separation. By modelling the mixing system as a matrix of FIR filters, the problems of stability of the inverse system are reduced when compared to using an AR system for example. Ideally if FIR filters were used to model the mixing system then an AR system would exactly represent the inverse separating system. However it is common to model the inverse system also as a FIR filter matrix due to stability. This has the implication that to ideally find the inverse of a mixing MA system, under the constraint that the inverse also has to be a MA system, the inverse FIR filter has to be infinitely long. An accurate architecture to invert a M^{th} order filter is an infinitive impulse response (IIR) filter of M^{th} order. However, stable IIR filters are limited to poles inside the unit circle and therefore, a stable IIR filter exists only for a minimum phase mixing system. FIR filters may be used to approximate the inverse solution. In real applications the mixing system may very well be a non-minimum phase system. One of the main problems with

using MA systems to model both the mixing and separating systems is that in a real environment the FIR filters require thousands of taps to accurately model the channel characteristics to a certain approximation. This introduces more computation, which reduces the convergence speed of BSS models. This is the tradeoff however to ensure stability of the demixing system which is necessary for real time applications. The M observed signals $\mathbf{x}(t)$ are coupled to the N reconstructed signals $\hat{\mathbf{s}}(t)$ via the demixing system. The demixing system has a similar structure to the mixing system. It contains $N \times M$ FIR filters of length Q , where $Q \geq P$. The demixing system can also be expressed as an $N \times M$ matrix $\mathbf{W}(t)$, with its element $\mathbf{w}_{ij}(t)$ being the impulse response from j th measurement to i th output. The reconstructed signal can be obtained as

$$\hat{\mathbf{s}}(t) = \sum_{\tau=0}^{Q-1} \mathbf{W}(\tau) \mathbf{x}(t - \tau). \quad (2.8)$$

2.5 Measuring Performance

There are numerous methods in the field of blind signal separation used to evaluate the performance of various separation algorithms including plots of separated signals, plots of cascaded mixing/demixing impulse responses and signal to noise ratios. It is important to use standard data test sets that are available to provide a unified methodology to making a good comparative analysis between algorithms in an objective manner. Controllable synthetic test cases are used to examine algorithm performance in trivial to moderately complex test cases allowing accurate evaluation of separation for different algorithms where information of sources and mixing/demixing systems are available. In comparison, however; it is also equally important to test the algorithms in a real environment to demonstrate the effectiveness of the algorithm in an application sense. Real world recordings for acoustic signal separation should be considered to reflect the complexity of real mixing systems and

the success of separation for different BSS algorithms. Sources, mixing systems and performance measures for synthetic and real cases that are standard tools for evaluating blind signal separation are referred to in (D. Schobben, K. Torkkola, and P. Smaragdis, 1999).

Despite all the different ways of formulating a method of separation, they are all related to a fundamental measure of statistical independence that can be used to derive differing objective and contrast functions. This is known as the Kullback-Leibler distance (KLD) measure. The Kullback-Liebler measures the distance between the joint probability distribution and the product of the marginal distributions as shown in Equation (2.9).

$$KL(p(y), \prod_{i=1}^n p(y_i)) = \int_{R^n} p(y) \log \frac{p(y)}{\prod_{i=1}^n p(y_i)} dy. \quad (2.9)$$

When

$$KL(p(y), \prod_{i=1}^n p(y_i)) = 0, \quad (2.10)$$

then the signals are independent otherwise the measure will be greater than zero. The idea behind BSS is to minimize the KLD. Comon used this measure and provided a unifying framework for Maximum Likelihood (ML), Infomax and Mutual Information (MI). Strictly speaking, the KLD is not a distance as it is asymmetric. ML and MI will be examined briefly along with other methods of deriving BSS objective functions to give a brief outline of the different approaches for solving BSS problems, but more focus will be given to methods that exploit non-stationary sources via joint-diagonalization.

As previously mentioned, there are numerous ways to evaluate the performance of a BSS algorithm. Lucas and Parra used the Signal to Interference Ratio (SIR) as a measure of the performance of the algorithm for their experimental results (Parra and Spence, 2000), as did Ikram and Morgan in (M. Ikram, and D. Morgan, 2000). In the

frequency domain the SIR is defined for a signal $s(t)$ in a multi-path channel $\mathbf{H}(\omega)$ at frequency bin ω to be the total signal powers of the direct channel versus the signal power stemming from the cross channels,

$$SIR[\mathbf{H}, \mathbf{s}] = \frac{\sum_{\omega} \sum_i |\mathbf{H}_{ii}(\omega)|^2 \langle |\mathbf{s}_i(\omega)|^2 \rangle}{\sum_{\omega} \sum_i \sum_{j \neq i} |\mathbf{H}_{ij}(\omega)|^2 \langle |\mathbf{s}_j(\omega)|^2 \rangle}. \quad (2.11)$$

In the case of known channels and source signals we can compute the expressions directly by using a sample average over the available signal and multiplying the powers with the given direct and cross channel responses. Replacing $\mathbf{H}(\omega)$ by $\mathbf{W}(\omega)\mathbf{H}(\omega)$, where $\mathbf{W}(\omega)$ is the inverse of $\mathbf{H}(\omega)$, defines SIR_O which is the output SIR. The objective of BSS methods is to obtain a high SIR improvement given by the ratio SIR_O/SIR_I .

Ikram and Morgan also define an SIR measure to assess the separation achieved for each source signal. In their experiment they implemented a TITO system in a real acoustic environment. For example, for source 1, the input SIR is given by,

$$SIR_{I,1} = \frac{\sum_{\omega} |\mathbf{H}_{11}(\omega)|^2 |\mathbf{s}_1(\omega)|^2}{\sum_{\omega} |\mathbf{H}_{12}(\omega)|^2 |\mathbf{s}_2(\omega)|^2}. \quad (2.12)$$

The SIR relations for source 2 are defined in a similar manner. Araki, Makino, Nishikawa and Saruwatari (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001) used a similar method of measuring the performance of a BSS algorithm. In their paper the noise reduction rate (NRR) was used and is defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB. The SNR is defined similarly as above.

$$NRR_i = SNR_{O,i} - SNR_{I,i}, \quad (2.13)$$

$$SNR_{O,i} = 10 \log \frac{\sum_{\omega} |\mathbf{A}_{ii}(\omega) \mathbf{s}_i(\omega)|^2}{\sum_{\omega} |\mathbf{A}_{ij}(\omega) \mathbf{s}_j(\omega)|^2}, \quad (2.14)$$

$$SNR_{I,i} = 10 \log \frac{\sum_{\omega} |\mathbf{H}_{ii}(\omega) \mathbf{s}_i(\omega)|^2}{\sum_{\omega} |\mathbf{H}_{ij}(\omega) \mathbf{s}_j(\omega)|^2}, \quad (2.15)$$

where $\mathbf{A}(\omega) = \mathbf{W}(\omega)\mathbf{H}(\omega)$ and $i \neq j$.

2.6 Stationary Instantaneous BSS

Many BSS algorithms for source signals that have been mixed in a linear, instantaneous fashion have been developed over the last decade. A brief explanation of some of the common methods to deriving objective functions and various methods of separation will be investigated.

BSS models that assume linear, instantaneous mixing are probably the simplest. Most of the algorithms that are developed for instantaneous BSS come from the field of Independent Component Analysis (ICA). ICA was originally developed to deal with problems relating to the *cocktail-party* problem. ICA is very closely linked to BSS in an instantaneous mixing climate. Like Principal Component Analysis (PCA), it is a linear transformation of multidimensional data to a new coordinate system. However where PCA only looks at decorrelation, ICA takes the process one step further by finding statistical independence. The ICA or instantaneous BSS problem is solved on the basis of minimizing or maximizing certain contrast or objective functions. Essentially the ICA problem is transformed into a numerical optimization problem. A good starting point to this section is to describe the various techniques for deriving the objective functions. A more detailed analysis will be provided for specific BSS algorithms that the proposed research presented has extended upon in following chapters.

The key to the ICA model is nonlinear decorrelation and/or nongaussianity, which are two kinds of higher-order information. Assuming that the column vector $\hat{\mathbf{s}}$ is the separated signals and the column vector \mathbf{x} is the mixture of the independent components then to estimate one of the components a matrix \mathbf{W} needs to be found. If \mathbf{W} were taken as a matrix that maximizes the nongaussianity of $\mathbf{W}^T \mathbf{x}$ then the independent components would be found up to an arbitrary scaling and permutation factor. The

idea of maximum nongaussianity is due to the central limit theorem, which states that sums of non-gaussian random variables are closer to Gaussian than the original ones. For an explanation of the central limit theorem refer to (A. Hyvärinen, J. Karhunen, and E. Oja, 2001). To use nongaussianity in ICA a quantitative measure needs to be defined. Such measurements for nongaussianity include kurtosis, which is a HOS while other general measurements for formulating objective functions stem from the general areas of Estimation theory, Information theory and PCA and include maximum likelihood, mutual information, negentropy and projection pursuit. Although there are numerous algorithms that slightly vary certain parameters, these are the general methods used to formulate the optimization problems required to determine the separation system \mathbf{W} , as in most cases a closed form solution does not exist.

2.6.1 Higher Order Statistics

Kurtosis is the most classical form of measuring nongaussianity and is used as an optimization criterion for the ICA problem. It is a fourth-order cumulant, which is a HOS. The kurtosis of a random variable \mathbf{y} , is defined as,

$$kurt(\mathbf{y}) = E\{\mathbf{y}^4\} - 3(E\{\mathbf{y}^2\})^2. \quad (2.16)$$

As a preprocessing function to simplify things the random variable \mathbf{y} is assumed to be centered (zero-mean), and have unit variance. Having unit variance, the right hand side (R.H.S) of Equation (2.16) reduces to $E\{\mathbf{y}^4\} - 3$. If the variable \mathbf{y} is Gaussian then $E\{\mathbf{y}^4\}$, which is the fourth moment, becomes $3(E\{\mathbf{y}^2\})^2$ and assuming unit variance then the R.H.S reduces to zero. So for a Gaussian random variable the kurtosis is zero while for a non-gaussian variable the kurtosis is non-zero. Kurtosis can be either negative or positive. If it is negative the random variable is said to be sub-gaussian while if it is positive the random variable is said to be super-gaussian. Each distribution can be also described by its *peakedness*; i.e. essentially the flatter

the distribution the more sub-gaussian the variable is, while spiky distributions such as the Laplace distribution describe super-gaussian variables. Many real-world data sets have super-gaussian distributions. The ICA estimation principle for separation is to maximise the nongaussianity, which in this case is to maximise a supergaussian distribution or minimise a subgaussian distribution via kurtosis. One of the main disadvantages of using kurtosis is that it is very sensitive to outliers and hence is not a robust measure of nongaussianity. Also for reliable estimation of HOS (cumulants and moments), it requires much more samples than for SOS.

2.6.2 Estimation Theory

Estimation theory is the study of trying to estimate a quantity of interest from a set of uncertain measurements. The type of estimation method will depend ultimately on the assumed data model. The data model of interest for this research is ICA. Typically a set of measurements will be used to estimate a set of parameters that describe the data. For example two parameters that are often needed are the mean μ and variance σ^2 . There are two ways of estimating a parameter set from available data. They are batch type estimation, which is also referred to as *off-line* estimation. For this all the measurement data must be firstly available. *On-line* estimation is the other technique where the estimates of the parameters are updated using incoming samples and recursion. This is also referred to as adaptive estimation. A good way to measure estimation errors is to use an error criterion such as the mean-square error (MSE) given as,

$$\varepsilon_{MSE} = E\{\|\theta - \hat{\theta}\|^2\}, \quad (2.17)$$

where θ and $\hat{\theta}$ are the true and estimated parameter vectors respectively. However caution is required in the way the parameters are estimated from the data, as some methods are not robust in that outliers adversely affect the estimated parameters. Other estimators include linear least squares method, but one of the main estimating

models used for ICA is the maximum likelihood (ML) method.

2.6.2.1 Maximum Likelihood

The ML method is based on the relatively simple idea that different probability models generate different samples and that any given sample is more likely to have come from some probability models than from others (J. Mendal, 1990). It is closely related to the infomax principle in the context of BSS. The ML method makes the assumption that the unknown parameters θ are constants and no prior information is available to describe them. The ML estimator corresponds to the value $\hat{\theta}_{ML}$ that makes the measured data most likely. Given the data from an experiment and an assumed model, ML determines the values for the parameters of the model, which most probably lead to the observed data and hence uses conditional probability. The likelihood equation is given as,

$$p(\mathbf{x}_T|\theta) = p(\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)|\theta). \quad (2.18)$$

Many density functions contain an exponential function, which increases the complexity of dealing with such models. It is often easier to work with the logarithm of the likelihood function, the loglikelihood function. The ML estimator is usually found from solving the loglikelihood function given as,

$$\frac{\delta}{\delta\theta} \ln p(\mathbf{x}|\theta)|_{\theta=\theta_{ML}} = 0. \quad (2.19)$$

One of the simplifying assumptions that is critical to ICA/BSS and allows ML to be used as a way of modeling the objective (contrast) function to be optimized is that of statistical independence. With this assumption the likelihood function decouples into the product

$$p(\mathbf{x}_T|\theta) = \prod_{j=1}^T p(x(j)|\theta) \quad (2.20)$$

Examples of BSS algorithms that use ML cost functions are given in (M. Feder, and E. Weinstein, 1988; L. Parra, C. Spence, and B. Vries, 1997; J. Cardoso, 1998).

These sections have briefly looked at the main ways of using estimation theory to derive objective functions for use in ICA. All the estimation methods mentioned above assume that the parameters θ are unknown deterministic constants. Bayesian estimation assumes that the parameters θ are random. These random parameters are modelled using a *priori* probability density that is known. One example of Bayesian estimation is Maximum a posteriori (MAP) estimation, which will not be explained but is worth mentioning for completeness.

2.6.3 Information Theory

Estimation theory for ICA is basically built on deriving a parametric model that provides the best estimate from which the observed data is obtained. In ICA, information theory is the other principal approach to formulating objective functions to be optimized. With information theory coding of the observed data is required. An understanding of entropy is required for information theory. Entropy is the fundamental concept of information theory. Entropy H is defined for a discrete-valued random variable \mathbf{X} as,

$$H(\mathbf{x}) = - \sum_i P(\mathbf{X} = a_i) \log P(\mathbf{X} = a_i), \quad (2.21)$$

where a_i are the possible values of \mathbf{X} . The entropy of a random variable is the degree of information that the observation of the variable gives. The more random, unstructured and unpredictable the variable is, the larger its entropy (J. Mendal, 1990). An important result relating to nongaussianity is that a Gaussian variable has the largest entropy among all random variables of equal variance. So entropy can be used as another measure of nongaussianity. Supergaussian distributions have very low entropy, as most of the values are concentrated in the same range of values. Negentropy, a form of differential entropy, provides a measure of nongaussianity that is always

positive and is zero for a Gaussian variable. It is defined as,

$$J(\mathbf{x}) = H(\mathbf{x}_{gauss}) - H(\mathbf{x}). \quad (2.22)$$

The only drawback of using this measure of nongaussianity is that it is computationally demanding. To avoid this ICA algorithms that are based on information-theoretic contrast functions use approximations to negentropy and also combine aspects of kurtosis and nonlinear transformations as previously defined. Approximations to negentropy are found using higher-order moments. An example of an approximation to negentropy is

$$J(\mathbf{x}) \approx \sum_{i=1}^p k_i [E\{G_i(\mathbf{x})\} - E\{G_i(v)\}]^2, \quad (2.23)$$

where k_i are positive constants, v is a standardized Gaussian variable, and G_i are nonquadratic nonpolynomial functions. This approach combines the robustness of negentropy however allows faster computation.

Another technique used to derive a contrast function for ICA is mutual information (MI). MI between n scalar random variables, \mathbf{x}_i for $i = 1, 2, \dots, n$ is defined as,

$$I(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{i=1}^n H(\mathbf{x}_i) - H(\mathbf{x}) \quad (2.24)$$

A good way of interpreting this result is to relate it to the concept of entropy and code length. The entropy term $H(\mathbf{x}_i)$ could be coded separately while $H(\mathbf{x})$ would provide the code length when \mathbf{x} is coded as a random vector, i.e. all the components are coded in the same code. MI shows the code length reduction of coding the whole vector $H(\mathbf{x})$ as opposed to coding the separate components individually. If the \mathbf{x}_i components are statistically independent then MI is zero otherwise it is non-negative. So if the MI used to model the ICA system were minimized in an optimization problem then this would infer that the components are as close to being statistically independent as possible thus providing near separation.

All of these techniques used to describe the separation of observed signals in an ICA framework are equivalent and can be linked through the KLD as shown in (J. Cardoso, 1998). It can be realized that ML and MI methods are just different realizations of Kullback-based contrast functions. As mentioned before after a particular contrast function has been chosen, various learning rules of optimization theory including natural and relative gradient descent introduced in (S. Amari, 1998), stochastic gradient descent, "Newton-like" algorithms and Lagrangian methods for constrained optimization, using both types of learning paradigms (i.e. batch learning and on-line learning), can be implemented to determine the optimum separating system that will provide the maximum separation of the independent components in the observed data. Another aspect which will be discussed in Chapter 5 to be considered in the context of optimization is using global optimization to solve the objective functions without the need for additional constraints or criteria.

2.6.4 Global Scaling and Permutation Ambiguities

Signal separation is not exact as there are two main ambiguities inherent in the process. Consider the multichannel instantaneous mixing system \mathbf{A} and the instantaneous multichannel demixing system \mathbf{B} . If \mathbf{BA} is the global system, which ideally should equal the identity matrix \mathbf{I} , the first ambiguity is the fact that the complete global transformation system \mathbf{BA} need not equal the identity matrix, but can equal any diagonal matrix \mathbf{D} . This means that separation of independent components can only be achieved up to an arbitrary scaling factor. The reason is that both the input \mathbf{s} and mixing system \mathbf{A} being unknown quantities, any scalar multiplier in one of the sources \mathbf{s}_i can always be cancelled by dividing the corresponding column \mathbf{a}_i of \mathbf{A} by the same scalar α_i ,

$$\mathbf{x} = \sum_i \left(\frac{1}{\alpha_i} \mathbf{a}_i \right) (\mathbf{s}_i \alpha_i). \quad (2.25)$$

The arbitrary scaling constants can be chosen in a variety of ways. For example the model could normalize all the output signals to have unity variance, normalize the output signals to have the same energy as the corresponding sensor signals or scale the output signals so that they have the same amplitude as the signal component of the corresponding sensor signal.

The second ambiguity is that the complete transformation \mathbf{BA} need not equal a diagonal matrix, but can equal any permuted diagonal matrix (P. Comon, C. Jutten, and J. Herault, 1991). The permutation matrix can be denoted as \mathbf{P} . Therefore the order of independent components cannot be determined. Combining these two ambiguities together ICA/BSS can only separate such that $\mathbf{BA} = \mathbf{PD}$. These ambiguities are important especially the permutation problem as it arises when separating convolutively mixed signals in the frequency domain and is a major constraint leading to the need for additional assumptions and information for separation such as required in geometric beamforming (L. Parra and C. Alvino, 2002).

2.7 Stationary Convolutive BSS

Section 2.4 covered the basic description of BSS with convolutive mixing. As mentioned previously, BSS in a convolutive mixing environment is closely related to the fields of blind deconvolution/equalization, where there is only one observed signal which consists of an unknown source signal mixed with itself at different time delays due to finite propagation speed and multipath propagation being a result of reverberations of some obstacles. The key to development of many convolutive BSS algorithms is the close relationship BSS of convolutive mixtures has to the instantaneously mixed ICA problems for MIMO systems.

To utilize the development of some of the common models for instantaneous ICA

problems, that were previously highlighted, the convolutive BSS problem needs to be transformed into a model consistent with ICA. Time domain methods to solving the convolutive BSS problem in particularly long reverberant environments are problematic for the reasons of poor convergence times due to more computations as outlined by (K. Torkkola, 1999; F. Ehlers, H. Schuster, 1997; L. Parra and C. Spence, 2000). The more adopted technique of solving this particular problem where there is a high number of unknown variables to solve for, is by transforming the convolutive time-domain model into the multiplicative frequency-domain model, which can benefit from the fast implementation of Fourier transforms and allows the problem to be modelled in a similar manner to its simpler instantaneous counterpart.

2.7.1 Frequency Domain Methods

Typically for the convolutive mixing BSS model described in Section 2.4, the time-domain convolutive problem is transformed into independent, multiple short-term or instantaneous mixing, BSS problems in the frequency domain via the T -point discrete Fourier transform (DFT). For the frame $[\mathbf{x}(t), \dots, \mathbf{x}(t + T)]$ this is given as $\mathbf{x}(\omega, t) = \sum_{\tau=0}^{T-1} e^{-i2\pi\omega\tau/T} \mathbf{x}(t + \tau)$. In a compact matrix-vector notation the time-frequency domain relationships are shown as

$$\mathbf{x}(t) = \mathbf{H}(t) * \mathbf{s}(t) \quad (2.26)$$

which in the frequency domain becomes,

$$\mathbf{x}(\omega) = \mathbf{H}(\omega)\mathbf{s}(\omega) \quad (2.27)$$

The reconstructed signals in the frequency domain are defined as

$$\hat{\mathbf{s}}(\omega) = \mathbf{W}(\omega)\mathbf{x}(\omega), \quad (2.28)$$

which corresponds to the original source signals, in the frequency domain, up to an arbitrary permutation and scaling factor; that is

$$\hat{\mathbf{s}}(\omega) = \mathbf{W}(\omega)\mathbf{H}(\omega)\mathbf{s}(\omega) = \mathbf{\Pi}(\omega)\mathbf{D}(\omega)\mathbf{s}(\omega). \quad (2.29)$$

In this case, $\mathbf{W}(\omega) \in \mathbb{C}^{N \times M}$ and $\mathbf{H}(\omega) \in \mathbb{C}^{M \times N}$. $\mathbf{D}(\omega) \in \mathbb{C}^{N \times N}$ is an arbitrary frequency dependent diagonal scaling matrix. The permutation matrix $\mathbf{\Pi}(\omega) \in \mathbb{R}^{N \times N}$ is frequency dependent and introduces frequency-dependent permutation errors in the output frequency response. In order to avoid the inherent frequency permutation problem it is desirable to either make $\mathbf{\Pi}$ independent of frequency or derive a criteria to ensure correct permutation alignment of all separated frequency bins.

2.7.2 Local Scaling and Permutation Ambiguities

The frequency permutation problem inherent in traditional time-frequency domain ICA models for convolutive mixing still places a restriction on the successful development of truly blind separation algorithms, (i.e. no additional assumptions or information). The problem is the assignment of signal contributions to different source channels consistently across different frequencies. In (L. Parra and C. Alvino, 2002), an attempt to resolve these ambiguities by adding prior information such as microphone position and the assumption that the sources are localized in space was given in what they have referred to as geometric source separation. Time domain algorithms for solving convolutive BSS problems are viable up to a certain number of dimensions. Depending on how reverberant the environment is, it is more viable to solve unknown demixing systems with a small to medium number of variables in the time domain than in the frequency domain. This is due to the extra computations required for firstly transforming the problem into the frequency domain via the DFT, and then solving the local frequency permutation problem. If we model the entire convolutive BSS problem in the time domain we eliminate the need to solve

the local frequency permutation problem. A more detailed analysis of a time domain BSS algorithm which illustrates this is given in Chapter 4. For systems with a higher number of variables the viability of transforming the problem into the frequency or subband domain becomes apparent. Combined with the limitations on the quality of separation for convolutive BSS problems in the frequency domain for environments with longer reverberation times, the advantages of using a subband model approach become evident. This will be shown in Chapter 6.

2.8 Non-stationary Sources

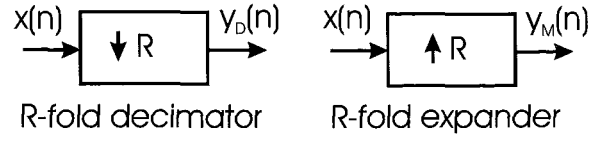
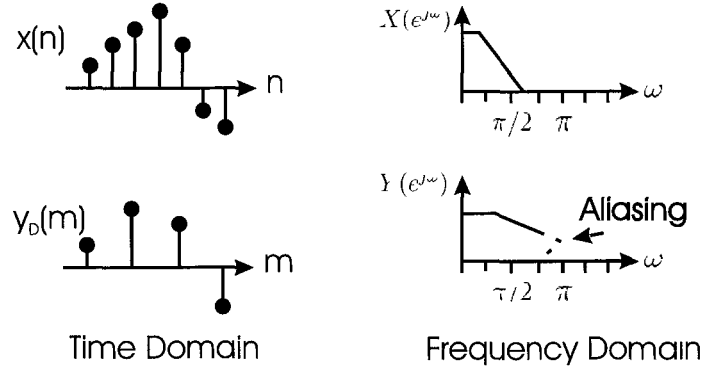
All the BSS algorithms discussed so far have made the assumption that the sources are i.i.d signals implying that they are stationary. One of the main reasons for computationally expensive algorithms when dealing with stationary sources is the fact that in addition to decorrelating or whitening the observed data using SOS, in order to achieve separation of the independent components a rotation matrix or orthogonal transformation needs to be applied using HOS such as cumulants or other nonlinear transformations. Using such HOS causes considerably longer convergence times for these algorithms. In many real cases the sources themselves often have varying statistical properties for example speech. As an alternative, any algorithms that model nonstationary sources can exploit time varying SOS using differential correlation instead of the much more cumbersome and laborious HOS approach. The merits and disadvantages of the previously discussed HOS were given as an initial investigation of how best to approach the problem of convolutive BSS. The focus of our proposed work is more closely related to problems that assume speech input, although the algorithms, models and concepts are applicable to a wide variety of applications. In this regard, we can make the natural assumption that the sources are non-stationary.

2.9 Subband Decomposition

The other main area that is vital to the research on extending BSS into the subband domain is multi-rate systems, filter banks and transform theory. Contributions by many researchers in the field of multi-rate systems have resulted in a mature theory of multi-rate systems. A good understanding of design methods, fundamental building blocks and the problems that arise when dealing with subband processing and reconstruction is necessary before BSS methods can be integrated. This section provides a brief overview of the current filter bank models that would be used for subband decomposition, the different components of the filter bank as well as the issue of perfect reconstruction (PR) and lossless systems. More efficient filter bank systems are developed using overlapped block transforms such as the Lapped Orthogonal Transform (LOT), the Modulated Lapped Transform (MLT) and the Extended Lapped Transform (ELT). The latter of these is probably the most viable for efficient and robust design but builds on the other two types of transforms mentioned and will be reviewed in Section 2.9.5.

2.9.1 Fundamental Concepts

In a multi-rate signal processing system there are four basic stages for probably the most simple direct structure filter bank design. Figure 1.2 showed an \tilde{M} -channel filter bank with these basic stages being the analysis filters, the R -fold decimator, the R -fold expander and the synthesis filters, which perform interpolation. The decimator is a device that reduces the sampling rate by an integer factor of R (down-sampling) while the expander is a device that increases the sampling rate by R (up-sampling). The use of different sampling rates offers benefits such as reduced computational complexity, which is ideal for the case of BSS as traditional cost functions suffer from long convergence times. Figure 2.2 shows the basic multi-rate building blocks.

**Figure 2.2** Multirate building blocks**Figure 2.3** Decimation by factor $R = 2$ in the time and frequency domain

Without going into too much mathematical detail an R -fold decimator takes an input sequence $x(n)$ and produces the output sequence

$$y_D(n) = x(nR) \quad (2.30)$$

where R is an integer. The effect of down sampling in the time and the frequency domain can be seen in Figure 2.3 with a factor of $R = 2$. In the frequency domain the result of decimating by a factor R is to expand the input spectrum by a factor of R . If the input spectrum has a bandwidth larger than π/R then the problem of aliasing occurs. When this occurs it is not possible to recover the original signal without a carefully designed low-pass or band-pass filter with bandwidth π/R .

An R -fold expander takes an input signal $y(m)$ and produces an output sequence $x(n)$ defined as follows,

$$x(n) = \begin{cases} y(n/R) & \text{if } n \bmod R = 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.31)$$

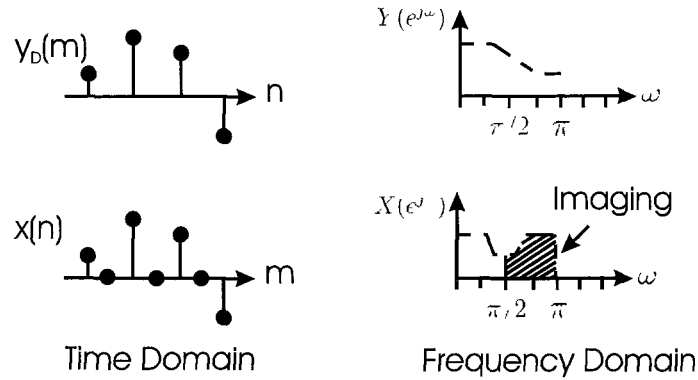


Figure 2.4 Expanding by factor $R = 2$ in the time and frequency domain

Inserting $R - 1$ zero samples between each pair of input samples forms the output sequence. Figure 2.4 shows the effect of up sampling by a factor of $R = 2$. In Figure 2.4 the output spectrum is formed by shrinking the original spectrum by a factor of R and replicating the shrunk spectrum $R - 1$ times. There is no aliasing however the replicas are due to the effect of imaging. To avoid this it is usual to follow the expander with a low-pass or band-pass filter having bandwidth equal to π/R , so that only one of the spectral images remains. Using a low-pass filter in the frequency domain has the effect of interpolation of samples in the time domain. Design of the analysis and synthesis filters will be reviewed in the following sections.

The analysis and the synthesis filters are Linear Time Invariant (LTI) systems, which means that the system is linear with respect to inputs and outputs, as well as having shift-invariance. For LTI systems, the system can be completely characterized by the impulse response $h(n)$ or in the z-domain, $\mathbf{H}(z)$. All rational transfer functions will have the following difference equation form

$$b_0 y(n) = - \sum_{m=1}^N b_m y(n-m) + \sum_{m=0}^N a_m u(n-m), \quad (2.32)$$

however, for this particular research more attention will be given to FIR filter implementation due to simplicity and stability. For the particular case of causal FIR filters,

for the analysis and synthesis filter banks, the R.H.S of Equation (2.32) reduces to the second term and in direct form the computational order is $N + 1$ multipliers, N adders, and N delays. There are two particular configurations for direct form structures of FIR filters and these are standard form and transposed form however for fast implementation a polyphase form is usually implemented. For more readings on how these particular configurations are derived refer to (P. Vaidyanathan, 1993; J. McClellan, R. Schafer, and M. Yoder, 1993).

2.9.2 Maximally Decimated Filter Banks and QMF

The fundamental parts of a typical subband signal processing system have been looked at. The analysis filter bank should successfully decompose the incoming signal into its various subband signals, and the synthesis filter bank ideally is able to recover a close approximation to the input signal from the subbands. Ideally this would mean that the reconstructed signal is simply a delayed version of the input signal as shown below,

$$\hat{x}(n) = x(n - D) \quad (2.33)$$

In order to make an educated choice in choosing a design for the analysis and synthesis filters it is a good idea to investigate the simplest design and extend upon that. Figure 1.2 presents probably the most simplest and typical subband processing system using the filter banks in the form Filter Bank Type 1 (FB-I). It is a \tilde{M} -channel maximally decimated QMF bank, which is uniform. If no subband processing is done then the computational complexity of running all the filters in the analysis filter bank has the same complexity as running a single filter without decimation. In the proposed research however there will be processing done on the subband signals in the form of BSS. In the search for perfect aliasing cancellation in design criteria, an initial review of QMF banks is needed.

Suppose that the QMF bank is the case where $\tilde{M} = 2$. Then a simple expression for the z-transform of the output signal is

$$\hat{X}(z) = \frac{1}{2}[H_0(z)F_0(z) + H_1(z)F_1(z)]X(z) + \frac{1}{2}[H_0(-z)F_0(z) + H_1(-z)F_1(z)]X(-z) \quad (2.34)$$

The aliasing component $X(-z)$ can be cancelled if the filters are chosen to meet the QMF condition

$$H_1(z) = H_0(-z), \quad F_0(z) = H_0(z), \quad F_1(z) = -H_0(-z). \quad (2.35)$$

This leads to,

$$\hat{X}(z) = \frac{1}{2}[H_0^2(z) - H_0^2(-z)]X(z), \quad (2.36)$$

and using Equation (2.33) the other criteria for perfect reconstruction is

$$H_0^2(z) - H_0^2(-z) = 2z^{-D}, \quad (2.37)$$

where D is some acceptable system delay. In practice, designing $H_0(z)$ to meet the alias cancellation condition is possible but only near approximation of the PR condition is met using nonlinear optimization techniques. The idea in QMF banks is to permit aliasing in the analysis bank instead of trying to avoid it and choose synthesis filters so that aliasing is cancelled.

The natural progression from studying the 2-channel QMF bank was the generalization to a QMF-like filter bank structure for R -channels. Pseudo-QMF filter banks were presented in (H. Nussbaumer, 1981). The basic idea behind this concept is that aliasing from adjacent bands is cancelled as the transition bandwidth of each one of the subband filters does not usually overlap with any other bands besides its neighbours. This can be seen in Figure 2.5. Ideally the different subbands would have infinite attenuation in the stopband, no transition band whatsoever and cutoff frequencies every π/\tilde{M} , however in practice this would require very high order FIR

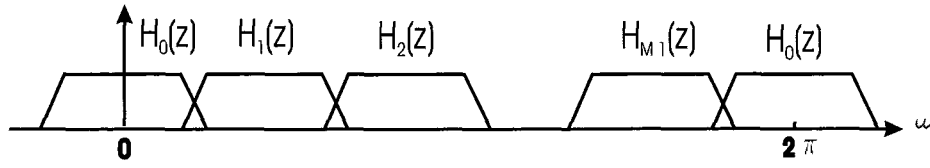


Figure 2.5 Frequency responses of an \tilde{M} th band filter $H(z)$

filters which is impractical for real time applications due to more computation and longer convergence times. As a result, neighbouring bands overlap thus causing aliasing as shown above.

The derivation of pseudo-QMF is left for the time being as is the polyphase implementation of such filter banks. For further detailed analysis on the proofs and the mathematics refer to (P. Vaidyanathan, 1993). The most common form that gives \tilde{M} -channel filter banks nearly perfect alias cancellation is where the synthesis filters are defined by

$$f_k(n) = h(n) \cos\left[\left(k + 1/2\right)\left(n - \frac{L-1}{2}\right)\frac{\pi}{\tilde{M}} + \phi_k\right], \quad (2.38)$$

for $k = 0, 1, \dots, \tilde{M} - 1$, and $n = 0, 1, \dots, L - 1$, where \tilde{M} is the number of subbands, L is the length of the filters, and the parameters ϕ_k control the relative phases of the modulating cosines. The analysis filters are obtained by time-reversing the synthesis filters

$$h_k(n) = f_k(L - 1 - n). \quad (2.39)$$

All the filters can be derived from a single filter $h(n)$ which is referred to as the *prototype*. In order to achieve aliasing cancellation by proper design the phases must be defined by

$$\phi_k = (-1)^k \frac{\pi}{4}, \quad (2.40)$$

or,

$$\phi_k = \left[1 + (-1)^k\right] \frac{\pi}{4}. \quad (2.41)$$

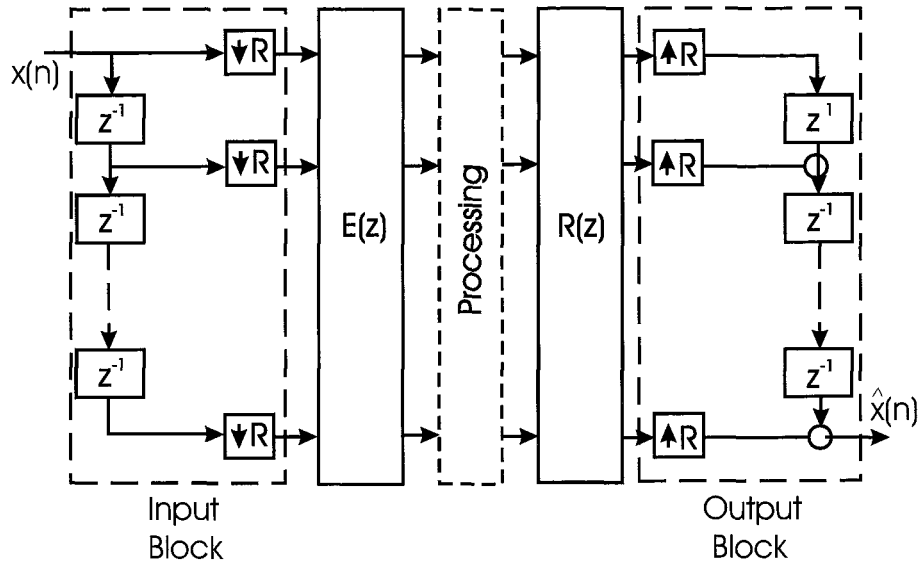


Figure 2.6 Polyphase structure of \tilde{M} -channel QMF maximally decimated filter bank

An understanding of these criteria is important when extending this design to block transforms and then lapped transforms, which will be explained in the following sections.

2.9.3 Block Transforms

Block transform theory is closely related to multi-rate systems and filter banks through structures based on the lossless polyphase designs of maximally decimated filter banks. Figure 1.2 showed a simplified structure of an \tilde{M} -channel maximally decimated filter bank. A more computationally efficient structure is the polyphase implementation. The polyphase structure, which is rearranged using the noble identities, is shown in Figure 2.6, which leads to PR. In order to obtain PR the following condition must hold:

$$R(z)E(z) = cz^{-m_0}I \quad (2.42)$$

where c is some constant and m_0 is some desired delay. This means that if

$$R(z) = \tilde{E}(z), \quad (2.43)$$

where,

$$\tilde{E}(z)E(z) = E(z)\tilde{E}(z) = I, \quad (2.44)$$

then PR will occur and $E(z)$ is paraunitary or lossless. $E(z)$ can also be orthogonal and is usually the case when developing block transforms. For a more detailed analytical analysis of deriving the polyphase structure shown and the concept of lossless systems an abundance of literature on multi-rate systems and filter bank theory is readily available.

Signal processing with a block transform is a special case of signal processing in subbands with a perfect reconstruction FIR filter bank (Malvar, 1992). The basic concepts of block transforms will be explained briefly before reviewing the basic motivations behind LOTs and ELTs, the latter of course being prevalent in the initial stages of the current research for subband decomposition.

To compute the transform of a signal $x(n)$, the signal must be partitioned into blocks. This can be shown as

$$x \equiv [x(m\tilde{M}), x(m\tilde{M} - 1), \dots, x(m\tilde{M} - \tilde{M} + 1)]^T \quad (2.45)$$

The dimension of x is \tilde{M} , which is referred to as the block size. Also a direct linear transformation matrix A on x can be given as,

$$X = A^T x, \quad (2.46)$$

and its inverse transform as,

$$x = [A^T]^{-1} X. \quad (2.47)$$

If the transformation matrix A is orthogonal than the inverse simply becomes,

$$A^T = A^{-1} \quad (2.48)$$

giving,

$$x = AX. \quad (2.49)$$

The columns of A are the basis functions of the transform. The number of basis functions in A is usually the same as the number of signal samples of x , which in this case is \tilde{M} . Each transformed block is the result of the inner product of the input block with its respective basis function in A .

The advantage of orthogonality is that the direct and inverse transformations are simply related by a transposition. Some examples of block transforms that have useful properties in areas of spectrum estimation, image coding, speech coding and adaptive filtering include the Discrete Fourier Transform (DFT), the Discrete Hartley Transform (DHT), the Karhunen-Loeve Transform (KLT), the Discrete Cosine Transform (DCT) and the Type-IV DCT.

With the paraunitary polyphase structure of the \tilde{M} -channel maximum decimated filter bank, some common properties of block transforms can be seen. The basis functions of the block transform correspond to the impulse responses of the synthesis filters, and the time-reversed basis functions would be the impulse responses of the analysis filters. Although the computational complexity of block transforms is good compared to PR FIR filter banks, one of the main restrictions of block-based transforms is that the FIR filter lengths (or basis function lengths) are of length \tilde{M} that causes poor stopband attenuation and leads to blocking effects in the reconstruction phase when quantising the transform coefficients. These blocking effects are a result of discontinuities at the boundaries of the blocks. For example, in audio applications a listener would here clicking at the block boundaries. Figure 2.7 gives a better understanding of the difference between the basis functions of block transforms and lapped transforms.

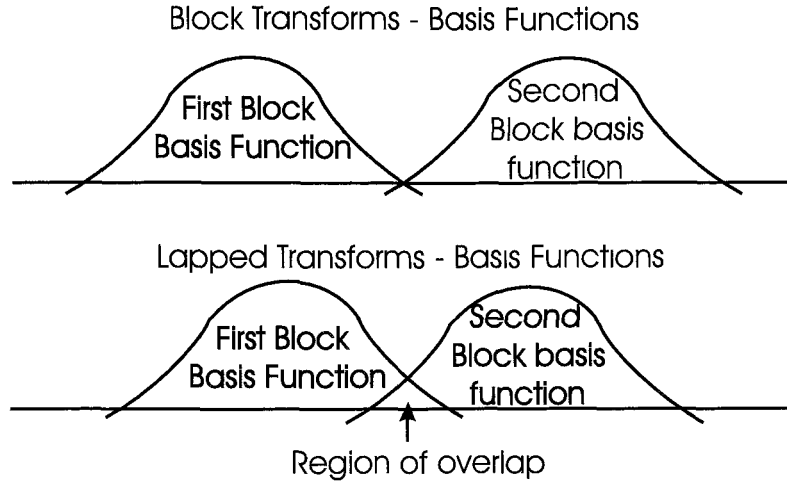


Figure 2.7 Basis functions of block transforms and lapped transforms

2.9.4 Lapped Transforms

The LOT has longer basis functions than traditional block transforms. In LOTs the basis function length is $2\tilde{M}$ where \tilde{M} is the number of subbands. The main reason for this is that if the basis functions are longer there will be a smoother transition to and from zero at the boundaries and the problem of blocking will be avoided. The following subsections will basically extend upon the idea of overlapping basis functions and finally arrive at the design of choice for subband decomposition in the current research, which is using a filter bank based on an ELT prototype function.

2.9.4.1 Lapped Orthogonal Transform

As briefly mentioned, a LOT unlike a block transform takes a signal and partitions it into blocks with $L = 2\tilde{M}$ samples in each block. These samples are transformed by an orthogonal matrix P of size $L \times \tilde{M}$ where \tilde{M} is the number of basis functions or transform coefficients. This means that there will be an overlap of $L - \tilde{M}$ samples when computing consecutive LT blocks. A more general definition of the length of the basis functions is $L = N\tilde{M}$. Block transforms are a special case of LOT where

$N = 1$.

Again the connection with multi-rate filter banks is made. A LOT is a filter bank in which the impulse responses of the synthesis filters are the LT basis functions and the impulse responses of the analysis filters are the time-reversed basis functions. This can be shown in the following equations.

$$f_k(n) = p_{nk} \quad (2.50)$$

where $k = 0, 1, \dots, N\tilde{M} - 1$ and $n = 0, 1, \dots, N\tilde{M} - 1$ and,

$$h_k(n) = f_k(N\tilde{M} - 1 - n) = p_{N\tilde{M}-1-n,k} \quad (2.51)$$

where p_{nk} is the element in the n th row and k th column of P .

2.9.4.2 Modulated Lapped Transform

Design techniques of how the filter responses for modulated lapped transforms are generated from a single prototype function $h(n)$ are similar to that of Pseudo-QMF outlined in Section 2.9.2. The MLT is similar to the LOT however, cosine modulation is used to generate all the different basis functions for both the analysis and synthesis filters. The normalized MLT basis functions can be written in the form

$$p_{nk} = h(n) \sqrt{\frac{2}{\tilde{M}}} \cos\left[\left(n + \frac{\tilde{M} + 1}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{\tilde{M}}\right] \quad (2.52)$$

This means that a careful design of the prototype function using appropriate windowing-techniques is all that is necessary to generate the responses of all the analysis and synthesis filters for the multirate filter bank. The only conditions on the design of the prototype is

$$h(L - 1 - n) = h(n), \quad (2.53)$$

and,

$$h^2(n) + h^2(n + \tilde{M}) = 1. \quad (2.54)$$

Due to the modulation structure of the MLT basis functions in Equation (2.52), very fast and efficient algorithms can be used to implement any filter banks that use these responses.

2.9.5 Extended Lapped Transform

Section 2.9.4.2 described how perfect reconstruction filter banks with a large number of subbands could be obtained from sinusoidal modulation of a single low-pass prototype and the correct choice of phase angles. In that section the length of the filter bank impulse responses or basis vectors were restricted to $L = 2\tilde{M}$. Lapped Transforms like the LOT and the MLT may not be adequate substitutes for QMF filter banks in applications where a strong subband separation is necessary. The usual lengths of QMF filters typically go from $4\tilde{M}$ to $16\tilde{M}$. It would be of practical value if the lengths of the basis functions could be made longer, without much penalty in the computational complexity (H. Malvar, 1992). This section reviews the scenario when perfect reconstruction is obtained for a modulated filter bank with a given number of subbands \tilde{M} and filter length $L > 2\tilde{M}$. With longer basis functions one can obtain better filtering performances. When there are basis functions of arbitrary length then this transform is referred to, as an Extended Lapped Transform (ELT) and the advantages of this particular transform is that better criteria for performance and computational complexity, in the same order of magnitude as standard block transforms, can be made. This section reviews the design of the prototype function $h(n)$ for an ELT and also reviews some of the fast algorithms that have been designed for implementation.

2.9.5.1 Basis Functions and the Prototype

All the basis functions or impulse responses of the different analysis and synthesis filters are defined using the same cosine modulation function that was used for the

MLT as shown in Equation (2.52). The phases of the different basis functions p_{nk} are related by

$$\phi_k = (k + \frac{1}{2})(N + \frac{1}{2})\frac{\pi}{2} \quad (2.55)$$

where $k = 0, 1, \dots, \tilde{M} - 1$, and $L = N\tilde{M}$. Another condition on the phases of the modulating cosines for PR is that

$$\phi_{k+1} - \phi_k = (2K + 1)\frac{\pi}{2} \quad (2.56)$$

From the above two equations it can be seen that a more general notation for the length of the basis functions is $L = 2K\tilde{M}$. K is referred to as the overlapping factor (H. Malvar, 1991). The length of the impulse responses of the FIR filters should be an even multiple of the number of subbands \tilde{M} . It should also be realized that for standard block transforms $K = 1/2$, for LOT and MLT $K = 1$, and for ELT $K \geq 2$.

To obtain PR with filters of any length, the prototype function $h(n)$ needs to satisfy a PR condition. The following equation is shown in scalar form but stems from the lossless condition mentioned in Section 2.9.3,

$$\sum_{\iota=0}^{2K-2s-1} h(n + \iota\tilde{M})h(n + \iota\tilde{M} + 2s\tilde{M}) = \delta(s) \quad (2.57)$$

where K is the overlapping factor and is greater than or equal to 2 for ELTs, $\iota = 0, 1, \dots, 2K - 1$, $s = 0, 1, \dots, K - 1$, $n = 0, 1, \dots, \tilde{M}/2 - 1$ and is the unitary impulse function. For $K = 2$, a parameterized family of windows (H. Malvar, 1991) can be obtained as

$$h(\frac{\tilde{M}}{2} - 1 - \iota) = -s_{\iota}s_{\tilde{M}-1-\iota}, \quad (2.58)$$

$$h(\frac{\tilde{M}}{2} + \iota) = s_{\iota}c_{\tilde{M}-1-\iota}, \quad (2.59)$$

$$h(\frac{3\tilde{M}}{2} - 1 - \iota) = c_{\iota}s_{\tilde{M}-1-\iota}, \quad (2.60)$$

$$h(\frac{3\tilde{M}}{2} + \iota) = c_{\iota}c_{\tilde{M}-1-\iota}, \quad (2.61)$$

where $c_i \equiv \cos(\theta_i)$ and $s_i \equiv \sin(\theta_i)$, for $i = 0, 1, \dots, \tilde{M}/2 - 1$, with the angles θ_i given by

$$\theta_i \left[\left(\frac{1-p}{2\tilde{M}} \right) (2i+1) + p \right] \frac{(2i+1)\pi}{8\tilde{M}} \quad (2.62)$$

where the free parameter p varies between $[0, 1]$ and controls the frequency responses of the filters in the analysis/synthesis banks.

PR conditions in the design of $h(n)$ do not have a unique solution as for any \tilde{M} and K there are an infinite number of solutions. A general optimal design technique for this is based on minimizing the stopband energy of the low-pass prototype defined by the objective function

$$E_s = \frac{1}{\pi} \int_{\omega_s}^{\pi} |H(e^{j\omega})|^2 d\omega \quad (2.63)$$

which is a constrained optimization problem where s defines the beginning of the stopband. A typical choice for ω_s is $1.2\pi/\tilde{M}$. Ideally the subband responses should be as close as possible to the ideal bandpass responses and so this forms the criteria for minimizing the stopband energy. A small value of E_s implies that there is virtually no aliasing among the subbands that are not neighbours to each other.

2.9.5.2 Performance of Fast Algorithms

One of the most beneficial characteristics of the ELT is that it can be efficiently computed, with a complexity close to that of standard block transforms without the problem of blocking effects. As previously mentioned Malvar derived some fast algorithms for the MLT and more importantly the ELT in the early 90's with good performance. These will be looked at briefly as further implementation will lean towards these more efficient structures in future research. Current research however will use the simple FIR filter bank structure shown in Figure 1.2 with the impulse responses of the analysis and synthesis filters being defined by the basis functions of the ELT transform.

The benefit of the ELT structure is that a good design for the prototype function $h(n)$ guarantees that all analysis and synthesis filters will have good bandpass responses. Also a structure of the modulating functions can lead to a regularly structured matrix, which leads to a fast algorithm. In (Malvar, 1992) Malvar presents a fast algorithm for any K using orthogonal butterfly angles and a type-IV DCT. The polyphase component matrix $E(z)$ is implemented as a cascade of two types of matrices: zero-delay orthogonal factors and pure delays. For a complete mathematical derivation of the algorithms for $K = 1$, $K = 2$ and $K \geq 2$ refer to the following (H. Malvar, 1992; H. Malvar, 1991; Malvar, 1992).

The final computational complexity of the fast ELT is given below

$$\frac{\tilde{M}}{2}(2K + \log_2 \tilde{M} + 3) \text{ multiplications} \quad (2.64)$$

$$\frac{\tilde{M}}{2}(2K + 3\log_2 \tilde{M} + 1) \text{ additions} \quad (2.65)$$

The computational complexity is close to that of block transforms, and thus is much lower than other filter banks such as the QMF bank, providing a motivation to use such a design for subband decomposition and BSS in the subband domain.

2.9.6 Conclusion

The ELT transform has been presented. The ELTs are a subclass of all PR filter banks in which the synthesis filters are identical, within time reversal, to the analysis filters as the orthogonality conditions are met. Filter lengths greater than twice the number of subbands can be used without much penalty in computational complexity and many subbands can be used. The benefits of using ELTs over other filter bank structures such as those based on the LOT, the MLT or the standard block transform have been reviewed and the implementation and structure of a fast implementation of a direct and inverse ELT algorithm has been briefly discussed.

2.10 Identifying Areas for Contribution

In general, the majority of literature for Blind Signal Separation focuses on assessing a proposed algorithm by evaluating and comparing the separation performance using some particular metric, as described previously, to a benchmark value. For acoustical applications that implement some convolutive BSS algorithm to perform separation of speakers in a room, the mixing model is usually set up with some pre-determined spatial layout. i.e. speakers and sensor array positions in a room are known. In (A. Westner, and V. Bove, 1999), a study of the nature of room impulse responses is conducted with discussion on obtaining room impulse responses taken from a conference room with different locations and source/sensor configurations. The details of the room layout can be seen in Figures 2.8 and 2.9. Figure 2.9 shows one particular configuration with two speakers and two microphones. The study of the nature of room impulse responses is given, with emphasis on frequency characteristics and artifacts of the specific room used for experimentation. Also the provision of acquiring ideal inverses of the room impulse responses is made using FIR matrix polynomial algebra techniques with reference to (R. Lambert, 1996). In convolutive mixing models, the impulse responses are usually modelled as FIR filters as mentioned previously with

an ideal inverse usually being an IIR filter. However, for purposes of stability, it is better to use a corresponding FIR filter that approximates the ideal inverse. Other work presented in this area for considering inversion of multivariate FIR MIMO systems by FIR MIMO systems can be found in (R. Rajagopal, 2000a). Additionally, in (M. Hofbauer, 2004a), the problem of Least-Squares optimal FIR inverse-filtering of a convolutive mixing system, given by acoustic impulse responses, is considered. The optimal filter is given by the LS-solution of a block-Toeplitz matrix equation, or equivalently by the time-domain multi-channel wiener filter. The latter of these two approaches is what we use in our framework in Chapter 3, for finding the optimal FIR MIMO inverse system of order Q , corresponding to the known FIR MIMO mixing system of the reverberant room. These approaches come from a more theoretical perspective, and use additional resources such as software to obtain room impulse responses which may not be available for reproducing the BSS experiment. In (A. Westner, and V. Bove, 1999), the acquisition of the impulse responses of the room are made using software that computes them approximately. The inversion process, as also described in (R. Lambert, 1996) is to apply standard scalar matrix algorithms to invert FIR polynomial matrices. In (A. Westner, and V. Bove, 1999), the following shows how to invert a 2×2 FIR matrix A .

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (2.66)$$

The inverse to A is:

$$W = A^{-1} = \frac{1}{a_{11} * a_{22} - a_{12} * a_{21}} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \quad (2.67)$$

In the above case, A is considered a square matrix, however for the non-square case the pseudo-inverse of the matrix is taken. i.e. $inv(A^H A) * A^H$. Generally for most test cases for evaluating new BSS algorithms, recordings or mixtures, made in controlled environments such as the one in (A. Westner, and V. Bove, 1999) are provided

as benchmark testing measures. A description of the physical layout of sensor positions with respect to the sources is usually also provided, and can be considered *a priori* information that can be used for good initialization for optimization procedures of objective functions if desired.

In an ideal testing case it would be better to have a framework which allows a simple method for obtaining all information for the mixing and demixing characteristics of the reverberant environment used in the experiment, in a *non-blind* fashion. Instead of only providing recordings of mixed signals using sensor arrays for benchmark testing with blind performance indicators, a framework for allowing knowledge of the full system would be more beneficial. This would not only provide a simple method for replicating experiments, but would allow a best possible solution using optimal filtering theory, that the proposed BSS solution could be compared to.

In Chapter 3 we address this issue by providing a simple framework or methodology for acquiring the full information of MIMO mixing/demixing systems in a reverberant environment, without the need for additional resources such as computer software for acquisition of room responses.

As stated previously one of the focal points of this thesis is the application of convolutive BSS problems with non-stationary sources such as speech in reverberant environments. In (L. Parra and C. Spence, 2000), a comprehensive analysis of the problem is provided starting with solving the simple instantaneous case first and then moving on to the more complex convolutive model. The pre-requisite knowledge of these mixing models was provided in Section 2.4. This particular work was a good starting point for the development of our own proposed BSS algorithms presented in Chapter 4. Algorithms presented in (L. Parra and C. Spence, 2000; L. Parra, C. Spence, and B. Vries, 1997), and (Parra and Spence, 2000), exploit the non-stationarity of acous-

tic sources. As explained in Section 2.8, there are two main approaches to solving the convolutive BSS problem. One is with the use of HOS, and one is the use of changing SOS in the temporal domain. Within the area of speech/acoustic applications, it is the latter method that is usually adopted when proposing a BSS model to be applied to an acoustics application.

In (L. Parra and C. Spence, 2000), a least-squares (LS) cost function is generated in the frequency domain to solve the convolutive BSS problem. The noise free criterion to optimize is

$$E(\omega, k) = \mathbf{W}(\omega) \bar{R}_x(\omega, k) \mathbf{W}^H(\omega) - \Lambda_s(\omega, k) \quad (2.68)$$

$$\begin{aligned} \hat{\mathbf{W}}, \hat{\Lambda}_s = \arg \min_{\mathbf{W}, \Lambda_s} \quad & \sum_{\omega=1}^T \sum_{k=1}^K \|E(\omega, k)\|^2 \\ & \mathbf{W}(\tau) = 0, \tau > Q \ll T, \end{aligned} \quad (2.69)$$

$$W_n(\omega) = 1$$

The method of joint diagonalization is used, however the inherent local frequency permutation problem that underpins all convolutive BSS problems is evident and deserves special attention. As described in Section 2.7.2, only consistent permutations for all frequencies will correctly reconstruct the unknown input signals. The solution proposed in (L. Parra and C. Spence, 2000) was to apply a fixed length FIR filter constraint. This constraint is performed by applying a projection operator P to the filter estimates every iteration. This requires transforming between time and frequency domains at each iteration which introduces computational overhead. In (M. Ikram, and D. Morgan, 2000), there is some evidence that the algorithm presented in (L. Parra and C. Spence, 2000) converges to a local minimum, resulting in insufficient cancellation of undesired cross signals from the convolutive mixture. Studies from (M. Ikram, and D. Morgan, 2000) prove that the ideas proposed in the existing literature are not capable of effectively handling the "permutation inconsistency" problem, which becomes worse as the length of room impulse response increase. Benefits of computational saving are evident for transforming to the frequency domain to perform multiplication in contrast to convolution in the time domain, for medium to longer unknown room responses. However, no extensive investigation in the literature has studied the comparison of a time domain approach using convolution for a smaller unknown system which avoids the local permutation problem, with the typical frequency domain approach with some constraint to solve the local frequency permutation problem.

In Chapter 4, with reference to (M. Ikram, and D. Morgan, 2000) and non-optimal performance of current methods such as those presented in (L. Parra and C. Spence, 2000), there is motivation to propose new BSS algorithms that solve the convolutive problem completely in the time domain for smaller systems and thus avoid the overhead of solving the inherent local frequency domain problem as well as the overhead

of transforming to and from other domains such as the frequency domain. In (S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, 2000; H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004a; M. Ikram, and D. Morgan, 2000), and (M. Ikram, and D. Morgan, 2002), emphasis is made on solving the permutation problem as opposed to avoiding it in the first place, by using knowledge of beamforming which assumes the distance between elements on the sensor array is known or the angle of incidence of the impinging waveform from sources to sensors is known. The merits of methods that avoid the local permutation problem rather than solve it have not been investigated for smaller to medium sized systems and provide motivation for work in this area. Two different algorithms based on avoiding the local permutation problem are proposed in Chapter 4, with initial emphasis on initialized gradient descent and Newton based methods of optimization.

Common frequency domain methods used for convolutive BSS problems with non-stationary sources such as speech, always have some particular criterion which requires some method of optimization. For convolutive BSS problems, transform to the frequency domain simplifies the separation process, however introduces a local permutation problem which must be resolved. The motivation for avoiding this local permutation problem leads to the work in Chapter 4. This work initially assumes a good initial starting point for the criterion in the region of the global minimizer. The justification of this is that the criterion in most cases defines a multivariate non-convex problem which if poorly initialized, will lead to a sub-optimal solution. Good initialization, requires additional *a priori* information, otherwise ill-convergence may result. Recently, much work has been done in the area of combining adaptive beamforming techniques and convolutive blind signal separation for the purposes of solving the inherent local permutation problem introduced. In (L. Parra and C. Alvino, 2002), the geometric information available to adaptive beamforming is exploited and

is introduced into the BSS algorithm as the initialization of filter parameters and as regularizations using penalty terms; both approaches are not limited to SOS approaches and can be extended to HOS.

When dealing with small to medium sized unknown demixing systems, instead of introducing beamforming techniques that either solve the permutation problem by analysing beamforming patterns to initialize unknown filters in the frequency domain before optimization or use constraint functions, a more direct method is to use the approach of global optimization. A good justification here is that geometric source separation techniques require additional resources to calculate beam patterns, or use knowledge of source or sensor array geometric configurations. In the true essence of BSS, when knowledge of such information is unknown, and the priority of solving or avoiding the local frequency domain approach is paramount, the combination of an uninitialized global optimization algorithm and a BSS convolutive method that avoids the local permutation problem, is adequately justified. Such a proposed method is introduced in Chapter 5.

In (L. Parra and C. Alvino, 2002), the use of geometric and adaptive beamforming with relation to BSS is explored. The motivation for using additional criteria is to avoid the local frequency permutation problem by utilizing additional information and to avoid ill-convergence to local minima. The work explored in (L. Parra and C. Alvino, 2002) will now be briefly elaborated on. Firstly with reference to beam patterns and the sensor array response, for a linear array with omnidirectional sensors and a far-field source, the sensor response depends only on the angle, $\theta = \theta(\mathbf{q})$, between the source and linear array,

$$\mathbf{d}(\omega, \mathbf{q}) = \mathbf{d}(\omega, \theta) = e^{-i(\mathbf{p}/c)\omega \sin(\theta)} \quad (2.70)$$

where $\mathbf{p} = [p_1, \dots, p_N]$ are the sensor positions of the array and c is the wave prop-

agation speed. In beamforming literature, each unknown FIR filter in the demixing system produces a beam or beam pattern. The quantity

$$\|r(\omega, \theta)\| = \|\mathbf{w}^H(\omega)\mathbf{d}(\omega, \theta)\| \quad (2.71)$$

can be plotted where the magnitude response of each beam pattern can be plotted as a function of frequency ω versus θ , the incident angle on the sensor array. From the plot we can visualize the frequency response of a given beam. With this information available a set of geometric constraints is then derived to solve the local permutation problem. An unconstrained objective BSS function used in (L. Parra and C. Spence, 2000), is used as the BSS criterion to be optimized as shown below,

$$J(\mathbf{W}) = \sum_{\mathbf{t}, \omega} \alpha(\omega) \|\mathbf{R}_{yy}(\mathbf{t}, \omega) - \text{diag}[\mathbf{R}_{yy}(\mathbf{t}, \omega)]\|^2. \quad (2.72)$$

The above criterion is a simultaneous diagonalization problem which, if unconstrained will suffer from ill-convergence and/or the local permutation problem. To avoid this, a geometric constraint is imposed. Firstly the assumption that the sources are localized at angles of $\theta = [\theta_1, \dots, \theta_M]$ is made. The response of the unknown FIR filters in demixing system \mathbf{W} for the directions in θ is given by $\mathbf{W}(\omega)\mathbf{D}(\omega, \theta)$, where $\mathbf{D}(\omega, \theta) = [\mathbf{d}(\omega, \theta_1), \dots, \mathbf{d}(\omega, \theta_M)]$. The penalty term added as a constraint to Equation (2.72) is,

$$J_{C1}(\omega) = \|\text{diag}[\mathbf{W}(\omega)\mathbf{D}(\omega, \theta)] - \mathbf{I}\|^2, \quad (2.73)$$

or

$$J_{C2}(\omega) = \|\mathbf{W}(\omega)\mathbf{D}(\omega, \theta) - \mathbf{I}\|^2. \quad (2.74)$$

As a summary of the work presented in (L. Parra and C. Alvino, 2002), the geometric information that either assumes knowledge of θ or the sensor positions \mathbf{p} in the linear array, was introduced into the convolutive BSS algorithm as initialization of the filter parameters to ensure reasonable convergence to a separating system that doesn't exhibit the local frequency permutation problem.

Similar work presented in (H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004a) proposes another way of estimating directions of arrival for solving the local permutation problem. The Moore-Penrose pseudoinverse $\mathbf{W}^+(f)$ of the separation matrix $\mathbf{W}(f)$ obtained by ICA is found. Then the direction θ_i of a source corresponding to the i -th row of $\mathbf{W}(f)$ is calculated by

$$\theta_i = \arccos \frac{\arg([\mathbf{W}^{-1}]_{ji}/[\mathbf{W}^{-1}]_{j'i})}{2\pi f c^{-1}(d_j - d_{j'})}, \quad (2.75)$$

where c is the propagation velocity and d_j is the position of sensor j . Again the assumption that a linear array of sensors is used with knowledge of the relative sensor positions d is made. This information may not always be available. Other works that adopt similar approaches with beamforming and assume priori knowledge of source localization for solving the local permutation problem include, (R. Aichner, and H. Buchner, 2003) and (M. Ikram, and D. Morgan, 2002).

Alternative to the methods discussed above, if information such as sensor locations or source localization angles are not known then poor initialization leads to poor separation performance, local permutation problems and ill-convergence. For small to medium sized systems where such information is not present, there is motivation to present a convolutive BSS algorithm that avoids the local permutation problem completely, as presented in Chapter 4, along with a global optimization algorithm that does not require any initialization information to find the correct separating system. Such a method is presented in Chapter 5.

So far we have discussed the motivation to propose a BSS solution to the convolutive BSS problem that avoids the permutation problem all together. The benefits of such a proposal would be seen systems that are small to medium sized. Also, a global optimization method is proposed for problems where information for good initialization is unavailable. For higher order demixing systems with longer impulse

responses that are reflective of highly reverberant environments, the motivation for maintaining the time domain convolutive BSS algorithm becomes less. The reason for this is that the computational overhead required for performing convolution in the time domain without needing to solve the local permutation problem becomes higher than the computational overhead of performing multiplication in the frequency domain and solving the local permutation problem introduced, as unknown dimensions increases.

The motivation for investigating convolutive BSS within a subband domain framework as opposed to a frequency domain framework does not come from a computational perspective, but rather from the limiting factor of separation performance by the frequency domain. In Chapter 6 we investigate BSS within a subband framework. In (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001) the limitation of frequency domain BSS is discussed by showing that the frequency domain BSS framework is equivalent to two sets of frequency domain adaptive beamformers. As a result it is shown that the performance of frequency domain BSS is upper bounded by that of adaptive beamformers. For a TITO system, two sets of beamformers are summarized as

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \quad (2.76)$$

An adaptive beamformer (ABF) is derived firstly for a null towards the jammer signal S_1 when S_2 is the target signal, and then secondly for a null towards the jammer S_2 when S_1 is the target signal. For a more precise derivation of how the equivalence is defined refer to (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001) and references therein. It is stated that although BSS and ABF can reduce reverberant sounds to some extent, they mainly remove sounds from the jammer direction which explains the poor separation performance of BSS in a room with longer reverberation

characteristics.

Also in (S. Araki, S. Makino, R. Aichner, T.Nishikawa and H. Sarawatari, 2003), a subband based blind source separation for convolutive mixtures of speech is investigated. The motivation for working in the subband domain comes from the draw-back of frequency domain BSS, i.e. when a long frame with a fixed frame shift is used to cover reverberation, the number of samples in each frequency becomes small and the separation performance is degraded. It is shown that unknown demixing FIR filters in each subband allow more effective separation with rooms with long reverberation. The subband model is presented as a series of stages. Firstly, a subband decomposition of the observed signals is realized using a single sideband modulation. Next a time-domain BSS algorithm is proposed based on time-delayed decorrelation for non-stationary signals. Finally, a synthesizing stage is used to combine separated subbands back into the fullband domain to obtain the recovered signals up to a global permutation and scaling factor.

Not much focus has been given to performing BSS within a subband framework. Obviously as discussed previously, methods to solve the inherent local permutation problem introduced by transforming convolution to multiplication in a different domain need to be proposed due to there being many uncoupled separation problems in the subband/frequency domain framework. In (S. Araki, S. Makino, R. Aichner, T.Nishikawa and H. Sarawatari, 2003), constraint null beamformers are used to initialize the unknown demixing FIR filters within each subband to prevent the local permutation problem. i.e. To design the initial value, an assumption that the mixing system $\mathbf{H} = \{h_{ji}\}$ represents only the time difference of sound arrival τ_{ij} with respect to the midpoint between the microphones is made. Then

$$\mathbf{H}(\omega) = \{h_{ji}(\omega)\} = \exp(j\omega\tau_{ji}), \quad (2.77)$$

where $\tau_{ji} = \frac{d_j}{c} \sin \theta_i$. Here an assumption is made that the position of the j -th microphone given by d_j is known, as well as the direction of the i -th source θ_i is known for the good initialization. For the work proposed in Chapter 6, a good investigation of a few different ways of designing the filter bank using extended lapped transforms is given for performing subband decomposition. Another issue to be addressed with regards to solving convolutive BSS problems within a subband domain framework is that no work has been proposed that jointly addresses the issue of getting improved separation performance, without needing additional assumptions on available information used for good initialization. This initialization is paramount in ensuring that the local permutation problem introduced by converting the problem from the time domain into some transform domain is addressed. If such information is not available in the problem space then methods that use global optimization need to be embraced and emphasis of combining good separation algorithms within subbands, with such optimization techniques is well justified.

Chapter 3

Acoustic Modelling

3.1 Introduction

The contribution of this chapter is to provide a repeatable general method to obtain all information of modelling a convolutively mixed BSS problem in a reverberant environment in a non-blind sense. In most cases, current literature provides recorded mixtures of speech or audio in a reverberant environment, or alternatively provides spatial layouts of rooms for setting up experiments with acoustic recordings and microphone arrays, as identified in Chapter 2. There is no obvious or direct method to determine the entire global system, comprised of the mixing and demixing systems. The benefit of having some model which allows total knowledge of the mixing/demixing systems for acoustic BSS mixing experiments and testing of BSS algorithms, is that an ideal solution generated from such a model can then be used as a benchmark for comparing the solution of the existing or newly proposed BSS algorithms to.

A better quantitative comparison of how well the proposed BSS algorithm performs can be made if we firstly assume a non-blind approach for that particular acoustic environment. Of course this only serves as a tool for conducting quality of separa-

tion of the algorithms and allowing a more effective comparative analysis tool for different BSS algorithms. Firstly we define a framework or methodology for obtaining impulse responses of a MIMO system so as to model the room response that couples each source to each sensor. By doing this we can have an analytical representation of the MIMO mixing system. After obtaining the information related to the mixing system, it is then necessary to obtain the corresponding inverting system that approximates a delayed identity MIMO system so that recovery of the initial sources is made up to an arbitrary system delay.

3.2 Acquiring Room Response

The focus of this work is on acoustic environments and in particular multi-channel room reverberation and finding a demixing system that will result in the separation of the unknown multiple nonstationary input sources from their mixtures. When evaluating the performance of separation algorithms in realistic scenarios where there is a high level of reverberation due to multipath propagation, it is often desirable to know the acoustical impulse response of actual rooms. A standard data set¹ and unified methodologies for testing separation performance of algorithms for synthetic and actual mixed signals measured in simulated and real reverberant environments are available from the authors of (D. Schobben, K. Torkkola, and P. Smaragdīs, 1999). Whilst most literature provides geometry of the rooms and experimental setup, as identified in Chapter 2, actual impulse responses for MIMO mixing systems of rooms are not usually provided. A common approach to obtain room impulse responses for is to use the maximum length sequence (MLS) method as described in (J. Borish, and J. Angell, 1983). With the MLS method it is possible to measure the acoustic im-

¹This data is available from [http //www2 ele tue nl/ica99/](http://www2.ele.tue.nl/ica99/)

pulse response with a great amount of accuracy and repeatability. The MLS method² is based on a cross-correlation technique and is highly immune to extraneous noise of all kinds. This property demonstrates usefulness for acoustical measurements in very noisy environments. The method uses a signal that is a deterministic periodic pseudo-random binary sequence and is used as a source signal to be propagated from source positions through loud speakers. It is generally required that its length be at least equal to the reverberation time of the room. It has similar spectral properties as true random white noise. From signal theory we know that the cross-correlation between an input signal $s(k)$ and an output signal $x(k)$ of a linear time-invariant (LTI) system, is related to the auto-correlation of the input by a convolution with the system impulse response,

$$R_{sx}(k) = R_{ss}(k) * h(k). \quad (3.1)$$

An important property of the signal used in the MLS method, like that of white noise, is that its auto-correlation function is approximately an impulse represented by the Dirac delta function,

$$R_{ss}(k) \approx \delta(k), \quad (3.2)$$

leading to

$$R_{sx}(k) = h(k). \quad (3.3)$$

Therefore the impulse response of the unknown channel can be found by cross-correlating the MLS input signal(s) $s(k)$ with the received/observed signal(s) $x(k)$ from the microphones. If using FIR filters to model the impulse responses of respective channels, then the length of the filters will depend on the sampling frequency and where the majority of information is contained from acoustic echoes, i.e direct sound and primary reflections from walls, floor, ceiling, and other objects.

²Generated with Aurora software plug-ins for CoolEditPro 2.1 available at <http://www.ramsete.com/Aurora/download/Aurora33Beta1.zip>

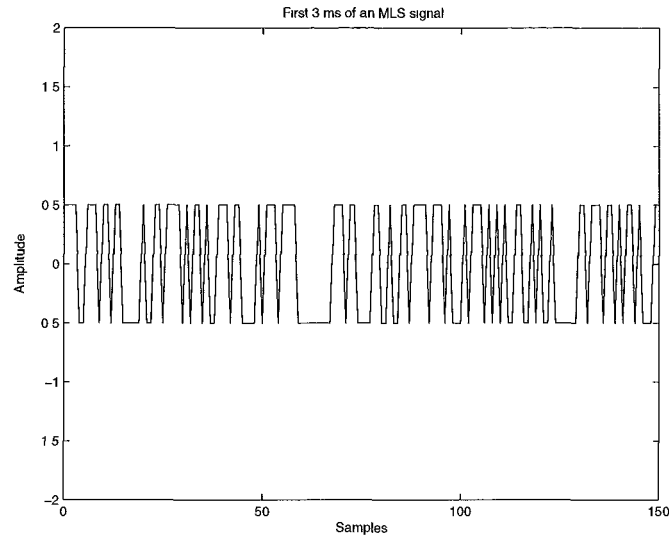


Figure 3.1 A section of a typical MLS signal

The most efficient technique to experiment with real world signals is to take a clean sound source such as that from the TIMIT corpus of speech and convolving it with a known impulse response of a room as can be obtained by the MLS method above for a MIMO system. By using artificially generated mixtures, we know what the mixing filters are and we can use them to determine how long separating filters need to be for good results. Additionally better quantitative analysis can be performed (A. Westner, and V. Bove, 1999). An example of a typical MLS signal is shown in Figure 3.1. To obtain the multi-channel impulse responses of a TITO mixing system in a reverberant room of dimension $2.28m \times 5.21m \times 3.45m$, the MLS signal(s) were played synchronously through two Genelec® 2029A loud-speakers. The mixed signals were recorded simultaneously at a sampling frequency of 48kHz using two Shure SM57 dynamic unidirectional microphones with a cardioid pickup pattern shown in Figure 3.2, which isolates the main sound source and minimizes background noise. All input/outputs were synchronized and interfaced to a Digi 001 PCI card through an 8 channel analogue I/O Digi 001 audio hardware box with 2 microphone preamps for recorded inputs available from the company Digidesign and is shown in Figure 3.3.



Figure 3.3 Digidesign Digi001 8 channel analogue I/O with 48kHz sampling.

The physical layout of the room is shown in Figure 3.4. Also for purposes of initialization for other typically used BSS methods to avoid ill-convergence and solve the permutation indeterminacy, the direction of arrival (DOA) angle is calculated using simple trigonometry. After calculating the TITO system impulse response at 48kHz, we downsample to 8kHz and discard the impulse response after a reverberation time of 200ms as the dominant echo information of the room is contained within. This corresponds to a FIR filter length of $P = 1600$ for each channel respectively. The resulting room impulse response coupling each source $s(t)$ to each sensor $x(t)$ is given

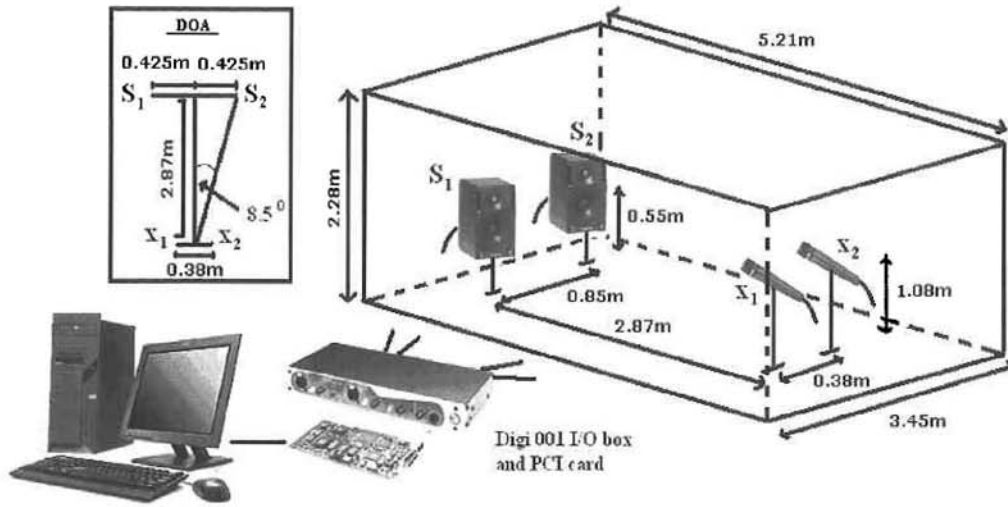


Figure 3.4 Geometrical properties of room layout and direction of arrival of sources impinging on array manifold.

in Figure 3.5. These results using this general non-blind approach to obtaining room mixtures are used for comparison with simulations that use a blind approach.

3.3 Inverting the Room

The second part of the contribution is to have a method to find the corresponding inverting MIMO demixing system to the MIMO room response found in the previous section. A process to identify the corresponding MIMO demixing system is necessary to provide an ideal solution to be used as a comparative benchmark for differing separation methods. This benchmark solution can be evaluated and compared against other solutions that are found using various BSS techniques, including the ones proposed in Chapter 4,5, and 6.

As described in Section 2.4, the MIMO demixing system $\mathbf{W}(t)$ required for deconvolution has a FIR structure which can be expressed in the z -domain using the causal

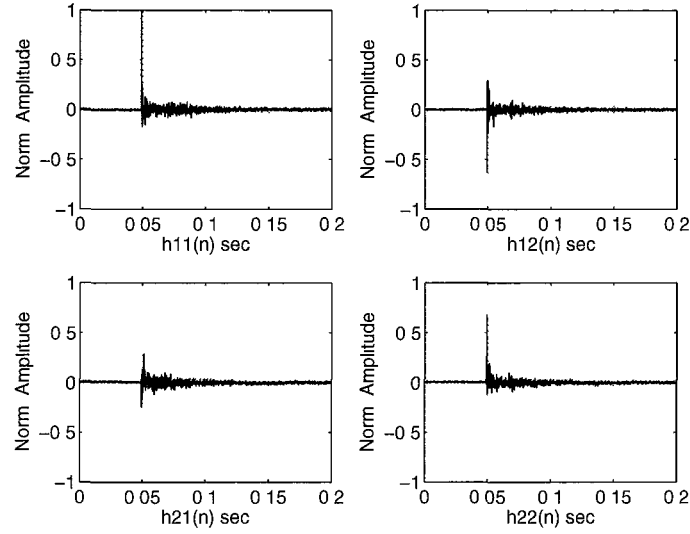


Figure 3.5 Measured room impulse responses with reverberation time of 200ms i.e. $P=1600$, for down-sampled rate of 8kHz for a TITO system.

z -transform

$$W_{ij}(z) = \sum_{n=0}^{\infty} w_{ij}(n)z^{-n}, \quad \forall i, \forall j \quad (3.4)$$

The demixing system is a matrix of FIR filters or moving average (MA), all-zero systems. Ideally, inversion of the known convolutive mixing MIMO system should be represented by an all-pole, auto-regressive (AR) system or IIR filters. However good approximation using inverse FIR filters is possible and provides a more stable process. The ideal inverse FIR filter will have an infinite length but this cannot be practically realized and so the inverse FIR filter channels must be approximated up to an arbitrary length that is suitable for the BSS application. A common assumption of BSS methods is that the order of the true channel(s) P or inverse channel(s) Q is known. In (A. Liavas, P. Regalia, and J. Delmas, 1999) effective channel order determination is made by applying a rank detection procedure to an over-modelled data covariance matrix. Other research that investigates finding the inversion of multivariate FIR MIMO systems by FIR MIMO systems and deriving conditions for the minimum FIR filter length is provided in (R. Rajagopal, 2000b). In our case, for

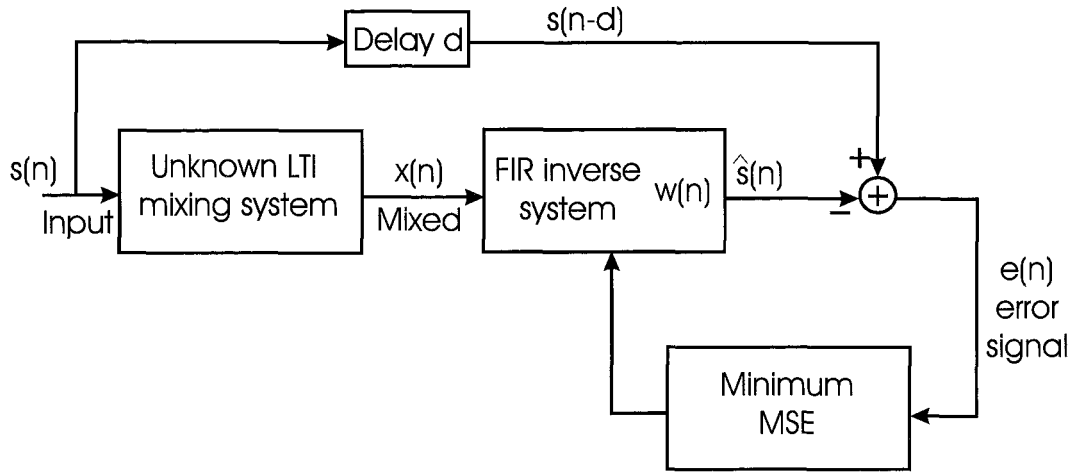


Figure 3.6 System identification of the inverse FIR filter for a SISO system.

applications of BSS to room acoustics, the channel order has been shown via experiment to be $P \approx 1600$ for 8kHz sampling frequency. To obtain a good estimate of the order of the demixing system Q for our BSS algorithm, the Wiener solution was calculated using optimal filtering theory, similar to that in (M. Hofbauer, 2004b).

The basic system identification model used to solve the discrete Wiener-Hopf normal equation for a single-input-single-output (SISO) system is given as a block diagram in Figure 3.6. According to the *orthogonality principle* given by Equation (5.94) in (A. Mertins, 1999), the following orthogonality condition for a SISO system must be satisfied by the optimal filter solution,

$$E\left\{[s(n-d) - \sum_{q=0}^{Q-1} w(q)x(n-q)]x^*(n-j)\right\} = 0, \quad (3.5)$$

where $j = 0, 1, \dots, Q-1$. For the MIMO FIR system case, this can be extended to the block diagram given in Figure 3.7. The aim is to find a set of NM filters \mathbf{w}_{ij} , $j = 1, \dots, M$, $i = 1, \dots, N$, with Q FIR coefficients, so that the output $\hat{s}_i(n)$ is an estimate of the desired signal, the delayed source, i.e. $s_i(n-d)$,

$$\hat{s}_i(n) = s_i(n-d) = \sum_j^M (\mathbf{w}_{ij}(n) * \mathbf{x}_j(n)). \quad (3.6)$$

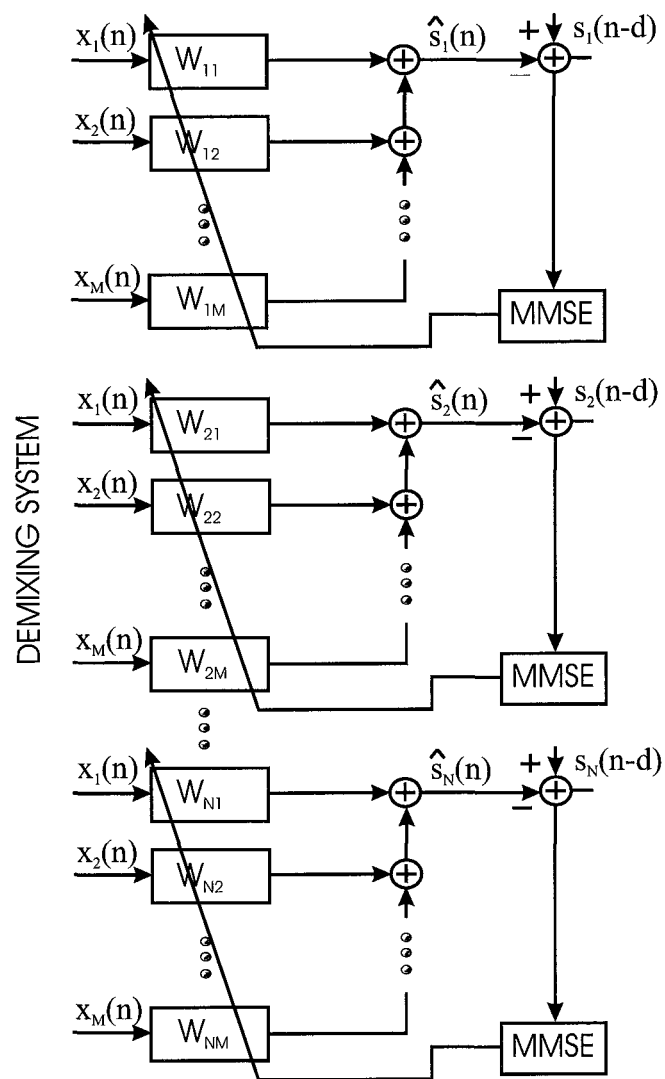


Figure 3.7 MIMO system identification for demixing FIR system using MMSE.

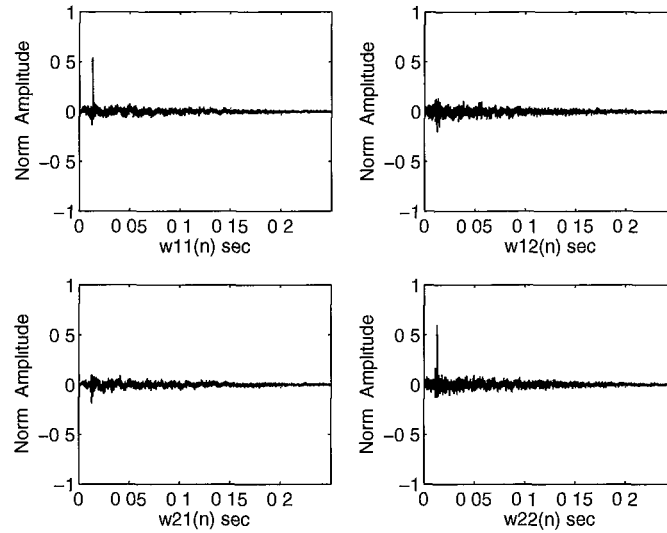


Figure 3.8 Wiener solution to TITO FIR demixing system with reverberation time of 250ms i.e. $Q=2000$, delay of for down-sampled rate of 8kHz.

In the MIMO case we are solving $N \times M$ Wiener-Hopf equations for a specified system delay d . Using segments of speech from the TIMIT corpus of speech and mixing them using the acquired mixing system given in Figure 3.5 we can obtain our mixtures $x(n)$ and desired signals $s(n - d)$ for some desired system delay d . Experimenting with different values of the order Q of the demixing system and a system delay d in milliseconds we use the Wiener solution for the MIMO FIR demixing system. Figure 3.8 gives the impulse responses for the TITO case which inverts the mixing process illustrated in Figure 3.5. The reverberation time is $T_R = 250\text{ms}$, corresponding to an order of the demixing system of $Q = 2000$ for a sampling frequency of 8kHz. The system delay that provided good results when looking at the global system shown in Figure 3.9, by cascading the mixing and demixing systems together, was $d = 50\text{ms}$.

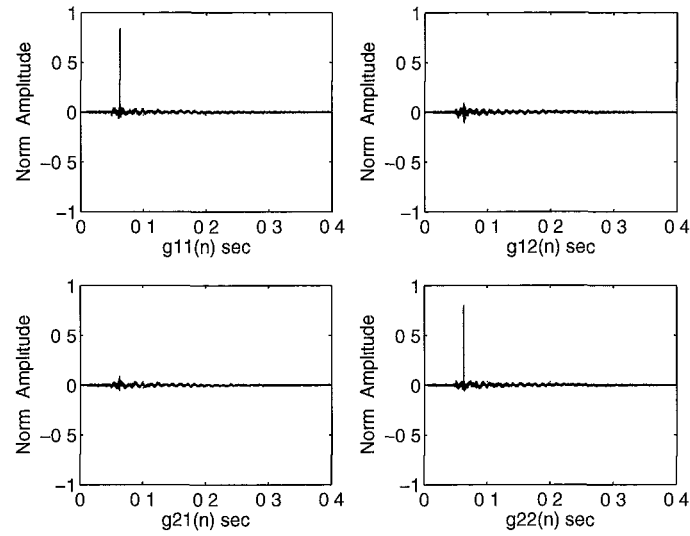


Figure 3.9 Global TITO FIR system, i.e. cascaded mixing and demixing FIR systems.

3.4 Conclusions

This chapter has provided a methodology for obtaining non-blind results for both a MIMO convolutive mixing system using FIR filters, and the corresponding MIMO convolutive demixing system of FIR filters that allows recovery of the input signals up to an arbitrary system delay. The importance of this is that this information about the unknown MIMO demixing system to be derived in the blind case is readily available and allows a better comparative analysis of how different BSS algorithms perform to our proposed BSS algorithms and with each other.

Chapter 4

Time and Frequency Domain Convolutional BSS models

4.1 Introduction

The two main contributions made in this chapter are two different algorithms that aim at solving the local frequency permutation problem that arises in convolutionally mixed BSS problems that transform a BSS model from the time to the frequency domain as is commonly the practice. The main reason for transformation to the frequency domain is the computational savings that are made by converting convolution in the time domain to multiplication in the frequency domain. Although this transformation benefits from the savings on computation just described, additional overhead is introduced from the transformation operator such as the DFT, as well as the computation introduced due to the introduction of the local frequency permutation problem which must be solved for successful separation up to an arbitrary delay and scaling factor.

The benefits of the computational savings of converting to a transform domain against the disadvantages of having to solve the local permutation problem within that domain must be assessed as there is a trade-off. By proposing two new BSS algo-

rithms that eliminate the local permutation problem by either staying completely in the time domain, or by performing common operations for all frequencies, we can assess the validity of performing the proposed approaches instead of the typically used frequency domain approach for various systems within the context of BSS in an acoustic environment. Computational complexity for convolutional mixing/demixing systems rises as the dimensions of the system increase. For large systems there is a large capacity to benefit from computational savings by converting to the frequency domain for example. The trade-off becomes good and when the overhead introduced from having to solve the local frequency permutation problem is minimal in comparison to the computational savings made by performing operations in the transform domain, the frequency domain method is justified. However, where a smaller mixing/demixing system exists, the trade-off of transforming to the frequency domain is not warranted and validates the proposed algorithms in this chapter. The motivation for both convolutional BSS algorithms in this chapter is to provide models that achieve separation up to a global permutation whilst avoiding the local frequency permutation problem which is commonly found in most convolutional BSS algorithms and must always be solved for, thus requiring additional computational overhead. The latter model will be integrated into our proposed subband domain model in Chapter 6 when dealing with larger more complex systems that are evident in reverberant environments.

The first algorithm presents an extension of the ACDC algorithm introduced in (A. Yeredor, 2002) for the instantaneous mixing problem to the more general convolutional mixing problem with nonwhite sources. Further assumptions made on the source signals are their mutual statistical independence, nonstationarity and smoothness of their power spectra. The algorithm iterates the estimation of the mixing system (AC step) and the source statistics (DC step) until convergence is achieved. The proposed

algorithm operates in the frequency domain, but unlike most frequency domain algorithms, it carries out some of the operations jointly for all frequencies. This allows us to overcome frequency dependent permutation and scaling problems.

In addition to the proposed ACDC algorithm, this chapter also proposes a new algorithm for solving the Blind Signal Separation (BSS) problem for convolutive mixing completely in the time domain. The closed form expressions used for first and second order optimization techniques derived in (M. Joho and K. Rahbar, 2002) for the instantaneous BSS case are extended to accommodate the more practical convolutive mixing scenario. Traditionally convolutive BSS problems are solved in the frequency domain (Rahbar and Reilly, 2003; Ikeda and Murata, 1999) but this requires additional solving of the inherent frequency permutation problem. Where this is good for higher order systems, systems with a low to medium number of variables benefit from not being subject to a transform such as the DFT. We demonstrate the performance of the algorithm using two optimization methods with a convolutive synthetic mixing system and real speech data. A summary of both algorithms is given in Section 4.4

4.2 ALS Approach

Consider a linear mixing process where N source signals $s_1(n), \dots, s_N(n)$ are mixed in a convolutive manner into $M \geq N$ observable signals $x_1(n), \dots, x_M(n)$. This operation can be expressed in matrix notation as

$$\mathbf{x}(n) = \sum_{m=0}^{P-1} \mathbf{H}(m) \mathbf{s}(n-m) \quad (4.1)$$

with $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$, $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_M(n)]^T$, and $[\mathbf{H}(n)]_{i,j} = h_{i,j}(n)$. The terms $h_{i,j}(n)$ denote the impulse responses from input j to output i of the mixing system. The aim is to identify the impulse responses $h_{i,j}(n)$

on the basis of the observed signals $x_1(n), x_2(n), \dots, x_M(n)$ up to the well known scaling and permutation ambiguities that are inherent to all blind identification and separation algorithms (J. Cardoso, 1998; Cichocki A. and Amari S.-I., 2002). The assumptions made about the input signals are their mutual statistical independence and quasi-stationarity over short periods of time, but nonstationarity in a more general sense, similar to (L. Parra and C. Spence, 2000; J. Anemuller, and B. Kollmeier, 2000; D. Pham, and J. Cardoso, 2001; K. Rahbar, J. Reilly, and J. Manton, 2002; A. Yeredor, 2002) and smoothness of their power spectra. These assumptions are for example well justified for independent speech signals. In particular, the assumption of nonstationarity of the sources allows us to solve the problem based on second-order instead of higher-order statistics (Cichocki A. and Amari S.-I., 2002; L. Parra and C. Spence, 2000; D. Pham, and J. Cardoso, 2001; A. Yeredor, 2002).

Approaches to solve the above mentioned blind identification problem can be divided into time and frequency domain methods. In this work, we use a frequency domain approach to transfer the convolutional time-domain mixing process into an instantaneous mixing processes in the frequency domain. Let $\mathbf{x}(\omega)$, $\mathbf{H}(\omega)$, and $\mathbf{s}(\omega)$ denote the Fourier transforms of the sequences $\mathbf{x}(n)$, $\mathbf{H}(n)$, and $\mathbf{s}(n)$, respectively. Assuming that a total number of K frequencies $\omega_k = 2\pi k/K$, $k = 0, 1, \dots, K-1$ are observed, we may replace Equation (4.1) with K instantaneous mixing processes of the form

$$\mathbf{x}(\omega_k) = \mathbf{H}(\omega_k)\mathbf{s}(\omega_k). \quad (4.2)$$

The aim is to find estimates $\hat{\mathbf{H}}(\omega_k)$ such that the remaining ambiguities can be expressed as

$$\hat{\mathbf{H}}(\omega) = \mathbf{H}(\omega)\mathbf{P}\mathbf{D}(\omega) \quad \forall \omega \quad (4.3)$$

where \mathbf{P} is a permutation matrix and $\mathbf{D}(\omega)$ is a (possibly frequency dependent) diagonal scaling matrix. Thus, we essentially look at a setting that is similar to the

one in (K. Rahbar, J. Reilly, and J. Manton, 2002). Differences between our approach and the one in (K. Rahbar, J. Reilly, and J. Manton, 2002) are that we assume colored instead of white source signals and that we use a different approach to determine the unknown mixing system. Our method can be seen as an extension of the ACDC algorithm that was introduced in (A. Yeredor, 2002) for instantaneous mixing. We transfer the ACDC algorithm to the frequency domain and optimize some of the involved parameters for all frequencies simultaneously. This joint optimization allows us to overcome the frequency dependent scaling and permutation ambiguity problems that occur with all frequency-domain approaches.

The notation for Section 4.2 is as follows. Vectors and matrices are printed in bold-face. The superscript $\{\cdot\}^H$ means transposition and complex conjugation of a matrix or vector. The superscript $\{\cdot\}^+$ denotes the pseudoinverse. $\mathbf{E}\{\cdot\}$ means the expectation operation. $\|\cdot\|_F$ is the Frobenius norm of a matrix. The symbols \otimes and \odot denote the Kronecker and Hadamard products, respectively. The term $\mathbf{v} = \text{diag}[\mathbf{u}]$ denotes the formation of a diagonal matrix \mathbf{v} from a set of values \mathbf{u} as well as forming a column vector \mathbf{v} from the diagonal elements of a matrix \mathbf{u} . If the argument is a set of matrices, then the result is a block diagonal matrix.

4.2.1 Algorithmic Model

We assume the signals $x_j(n)$ to be observed during T different time epochs and rewrite Equation (4.2) as

$$\mathbf{x}(\omega_k, t) = \mathbf{H}(\omega_k)\mathbf{s}(\omega_k, t), \quad t = 1, 2, \dots, T \quad (4.4)$$

Given the observations $\mathbf{x}(\omega_k, t)$ it is straightforward to find estimates for the cross-power spectral density matrices $\mathbf{R}_{\omega_k, t} = \mathbf{E}\{\mathbf{x}(\omega_k, t)\mathbf{x}^H(\omega_k, t)\}$, and based on Equation (4.4), these can be described as

$$\mathbf{R}_{\omega_k, t} = \mathbf{H}(\omega_k)\mathbf{\Lambda}_{\omega_k, t}\mathbf{H}^H(\omega_k), \quad (4.5)$$

where $\Lambda_{\omega_k, t}$ are the cross-power spectral density matrices of the nonstationary input processes $s(\omega_k, t)$. Because the different sources are assumed to be statistically independent, the matrices $\Lambda_{\omega_k, t}$ are diagonal.

The criterion to find estimates for the mixing systems $\mathbf{H}(\omega_k)$ and the unknown input power spectra $\Lambda_{\omega_k, t}$ is defined as

$$\min C = \sum_{k=0}^{K-1} C_{\omega_k}, \quad (4.6)$$

with

$$C_{\omega_k} = \sum_{t=1}^T \|\mathbf{R}_{\omega_k, t} - \mathbf{H}(\omega_k) \Lambda_{\omega_k, t} \mathbf{H}^H(\omega_k)\|_F^2. \quad (4.7)$$

The minimization involves two steps that are repeated until convergence. First, we carry out a so-called AC step, where the criterion is minimized with regard to individual columns of $\mathbf{H}(\omega_k)$, while matrices $\Lambda_{\omega_k, t}$ remain constant (cf. (A. Yeredor, 2002)). This step is carried out repeatedly for all columns of $\mathbf{H}(\omega_k)$, until convergence. In the second step, the DC step, C is minimized with respect to $\Lambda_{\omega_k, t}$. Then, another AC step is carried out, and so on, until the final minimum is reached. In order to minimize the effect of different scale factors and permutations at different frequencies, a projection procedure is included in the AC step that ensures that the identified time-domain impulse responses do not significantly exceed a maximum, pre-determined length.

4.2.1.1 The AC Step (Part 1)

In this step, we minimize C_{ω_k} with respect to the ℓ th column of $\mathbf{H}(\omega_k)$ for each frequency ω_k separately. Using the equality

$$\mathbf{H}(\omega_k) \Lambda_{\omega_k, t} \mathbf{H}^H(\omega_k) = \sum_{n=1}^N \lambda_n^{(\omega_k, t)} \mathbf{h}_{\omega_k, n} \mathbf{h}_{\omega_k, n}^H, \quad (4.8)$$

where $\mathbf{h}_{\omega_k, n}$ is the n th column of $\mathbf{H}(\omega_k)$, $\lambda_n^{(\omega_k, t)} = [\Lambda_{\omega_k, t}]_{n, n}$ are the diagonal elements of $\Lambda_{\omega_k, t}$, and

$$\mathbf{R}_{\omega_k, t}^{(\ell)} = \mathbf{R}_{\omega_k, t} - \sum_{n=1, n \neq \ell}^N \lambda_n^{(\omega_k, t)} \mathbf{h}_{\omega_k, n} \mathbf{h}_{\omega_k, n}^H, \quad (4.9)$$

we can write

$$C_{\omega_k} = \sum_{k=1}^K \|\mathbf{R}_{\omega_k, t}^{(\ell)} - \lambda_{\ell}^{(\omega_k, t)} \mathbf{h}_{\omega_k, \ell} \mathbf{h}_{\omega_k, \ell}^H\|_F^2. \quad (4.10)$$

Similar to (A. Yeredor, 2002) for instantaneous mixing, this criterion can be rewritten as

$$C_{\omega_k} = \tilde{C}_{\omega_k} - 2\mathbf{h}_{\omega_k, \ell}^H \mathbf{P}_{\omega_k, \ell} \mathbf{h}_{\omega_k, \ell} + p_{\omega_k, \ell} (\mathbf{h}_{\omega_k, \ell}^H \mathbf{h}_{\omega_k, \ell})^2, \quad (4.11)$$

with

$$\mathbf{P}_{\omega_k, \ell} = \frac{1}{2} \sum_{t=1}^T \lambda_{\ell}^{(\omega_k, t)} [\tilde{\mathbf{R}}_{\omega_k, t}^{(\ell)H} + \tilde{\mathbf{R}}_{\omega_k, t}^{(\ell)}], \quad (4.12)$$

and

$$p_{\omega_k, \ell} = \sum_{t=1}^T [\lambda_{\ell}^{(\omega_k, t)}]^2. \quad (4.13)$$

The optimal vector $\mathbf{h}_{\omega_k, \ell}$ is given by

$$\mathbf{h}_{\omega_k, \ell} = b_{\omega_k, \ell} \beta_{\omega_k, \ell} \quad (4.14)$$

where $\beta_{\omega_k, \ell}$ is the unit-norm eigenvector of $\mathbf{P}_{\omega_k, \ell}$ that corresponds to the largest positive eigenvalue $\mu_{\omega_k, \ell}$ (A. Yeredor, 2002). The optimal prefactors $b_{\omega_k, \ell}$, computed separately for each frequency, are given by $b_{\omega_k, \ell} = (\mu_{\omega_k, \ell} / p_{\omega_k, \ell})^{1/2}$.

4.2.1.2 The AC Step (Part 2)

Minimizing the objective criterion for each frequency separately does not allow us to resolve any of the permutation and scale ambiguities. Therefore, at this stage, we employ a projection technique that is similar to the one in (K. Rahbar, J. Reilly, and J. Manton, 2002) in order to jointly compute the prefactors $b_{\omega_k, \ell}$ that result in time domain responses $h_{i,j}(n)$ of given, arbitrary length P . The vectors $\beta_{\omega_k, \ell}$ remain

the ones computed in Part 1 of the AC step. Note that concentrating $h_{i,j}(n)$ in time will also yield smooth frequency responses $H_{i,j}(\omega)$, which are characteristic for most real-world mixing systems.

Let

$$\mathcal{B}_\ell(n) = [e^{j\omega_0 n} \beta_{\omega_0, \ell}, e^{j\omega_1 n} \beta_{\omega_1, \ell} \dots, e^{j\omega_{K-1} n} \beta_{\omega_{K-1}, \ell}]. \quad (4.15)$$

Then, the time-domain impulse responses $\mathbf{h}_\ell(n) = [h_{1,\ell}(n), \dots, h_{M,\ell}(n)]^T$ that correspond to $\mathbf{h}_{\omega_k, \ell}$ are given by

$$\mathbf{h}_\ell(n) = \frac{1}{K} \mathcal{B}_\ell(n) \alpha_\ell, \quad (4.16)$$

with $\alpha_\ell = [b_{\omega_0, \ell}, b_{\omega_1, \ell}, \dots, b_{\omega_{K-1}, \ell}]^T$. The vector α_ℓ that maximizes

$$\sum_{n=0}^{P-1} \mathbf{h}_\ell^H(n) \mathbf{h}_\ell(n) \text{ subject to } \sum_{n=0}^{K-1} \mathbf{h}_\ell^H(n) \mathbf{h}_\ell(n) = 1, \quad (4.17)$$

is the one that maximizes $\alpha_\ell^H \Psi \alpha_\ell$ subject to $\alpha_\ell^H \Omega \alpha_\ell = 1$ with

$$\Psi = \sum_{n=0}^{P-1} \mathcal{B}_\ell^H(n) \mathcal{B}_\ell(n), \quad \Omega = \sum_{n=0}^{K-1} \mathcal{B}_\ell^H(n) \mathcal{B}_\ell(n). \quad (4.18)$$

This optimal vector α_ℓ is given by the eigenvector that corresponds to the largest eigenvalue ρ of the generalized eigenvalue problem $\Psi \alpha_\ell = \rho \Omega \alpha_\ell$, normalized such that $\alpha_\ell^H \Omega \alpha_\ell = 1$.

4.2.1.3 The DC Step

In this step, we minimize C with respect to $\lambda_{\omega_k, t}$. We first rewrite the criterion for each t and frequency ω_k as a squared Euclidean norm of a difference vector (cf. (A. Yeredor, 2002))

$$C_{\omega_k, t} = \|\mathbf{a}_{\omega_k, t} - \mathcal{H}_{\omega_k} \lambda_{\omega_k, t}\|_2^2, \quad (4.19)$$

with

$$\begin{aligned} \lambda_{\omega_k, t} &= \text{diag}[\lambda_{\omega_k, t}], \\ \mathbf{a}_{\omega_k, t} &= \text{vec}(\mathbf{R}_{\omega_k, t}), \\ \mathcal{H}_{\omega_k} &= (\mathbf{H}^*(\omega_k) \otimes \mathbf{1}) \odot (\mathbf{1} \otimes \mathcal{H}(\omega_k)). \end{aligned} \quad (4.20)$$

Computing the optimal vectors $\lambda_{\omega_k,t}$ for each frequency separately is straightforward: $\lambda_{\omega_k,t} = \mathcal{H}_{\omega_k}^+ \mathbf{a}_{\omega_k,t}$. However, such an approach has the drawback that the large degree of freedom may make the tradeoff between the spectral properties of the sources and the mixing system too easy to allow for an accurate estimation of the true underlying random processes on the basis of a finite number of observations. Therefore, algorithms like the one in (K. Rahbar, J. Reilly, and J. Manton, 2002) simplify the source modelling to white sources and absorb all coloration into the mixing system. In practice, however, one often has some *a priori* knowledge about the source signals, which could be exploited during the blind identification process. In the following, we assume that the power density spectra $\lambda_n^{(\omega_k,t)}$ are smooth functions of frequency and that spectral samples $\lambda_n^{(\omega_k,t)}$, $k = 1, 2, \dots, K$ can be well approximated in the form

$$\lambda_n^{(t)} = \mathbf{B} \mathbf{v}_n^{(t)} \quad (4.21)$$

with $\lambda_n^{(t)} = [\lambda_n^{(\omega_0,t)}, \lambda_n^{(\omega_1,t)}, \dots, \lambda_n^{(\omega_{K-1},t)}]^T$, where \mathbf{B} is a $T \times Q$ matrix with $Q < T$ whose columns contain appropriate smooth basis functions.

We now consider the simultaneous optimization of all unknown values $\lambda_n^{(\omega_k,t)}$ for a given t using the approximation given by Equation (4.21). For this, we first define the cost function

$$C_t = \sum_{k=0}^{K-1} C_{\omega_k,t} = \|\mathbf{a}_t - \mathcal{H} \lambda_t\|_2^2, \quad (4.22)$$

where

$$\lambda_t = \begin{bmatrix} \lambda_{\omega_0,t} \\ \vdots \\ \lambda_{\omega_{K-1},t} \end{bmatrix}, \quad \mathbf{a}_t = \begin{bmatrix} \mathbf{a}_{\omega_0,t} \\ \vdots \\ \mathbf{a}_{\omega_{K-1},t} \end{bmatrix}, \quad (4.23)$$

$$\mathcal{H} = \text{diag}[\mathcal{H}_{\omega_0}, \mathcal{H}_{\omega_1}, \dots, \mathcal{H}_{\omega_{K-1}}]. \quad (4.24)$$

Equation (4.21) can be rewritten as

$$\lambda_t = \bar{\mathbf{B}} \mathbf{v}_t, \quad \bar{\mathbf{B}} = [\mathbf{B} \otimes \mathbf{I}_{N \times N}], \quad (4.25)$$

and hence we can write

$$C_t = \|\mathbf{a}_t - \bar{\mathcal{H}}\mathbf{v}_t\|_2^2, \quad (4.26)$$

with $\bar{\mathcal{H}} = \mathcal{H}\bar{\mathbf{B}}$. The vector \mathbf{v}_t that minimizes Equation (4.26) is given by

$$\mathbf{v}_t = \bar{\mathcal{H}}^+ \mathbf{a}_t. \quad (4.27)$$

Given \mathbf{v}_t , the vector λ_t , containing all values $\lambda_n^{(\omega_k, t)}$ required to set up the matrices $\lambda_{\omega_k, t}$ for the next AC step, are found from Equation (4.23). Depending on the basis \mathbf{B} used, it is not assured that all $\lambda_n^{(\omega_k, t)}$ turn out positive. Therefore, we include one more step where any negative values $\lambda_n^{(\omega_k, t)}$ are set to zero.

4.2.2 Simulation Results

We consider the case where two nonstationary, colored source signals are mixed with a two-input two-output mixing system. The autocorrelation sequences of the sources were randomly generated for each time epoch as $r_{ss}^{(i, t)}(m) = \sigma_{i, t}^2 c_{i, t}(m) * c_{i, t}(-m)$ where $\sigma_{i, t}^2$ are uniform random variables and $c_{i, t}(m)$ are length-3 sequences of real-valued Gaussian random variables. The mixing system was chosen as

$$\mathbf{H}(0) = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{H}(1) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (4.28)$$

This is a paraunitary system, and its inverse, delayed to become causal, is given by $\mathbf{G}(0) = \mathbf{H}(1)$, $\mathbf{G}(1) = \mathbf{H}(0)$. For the cascade of both systems we have $\mathbf{C}(n) = \sum_{m=0}^1 \mathbf{G}(m)\mathbf{H}(n-m) = \delta_{n,1}\mathbf{I}$.

In the tests, $K = 64$ frequency points were considered. The basis sequences included in the columns of matrix \mathbf{B} for the DC step were chosen as

$$B_{i, j} = \gamma_j \cos\left(\frac{2\pi i j}{T}\right), \quad \gamma_j = \begin{cases} \sqrt{2/T}, & j \in \{0, \frac{T}{2}\} \\ \sqrt{1/T}, & \text{otherwise} \end{cases} \quad (4.29)$$

with $j = 0, 1, \dots, J$ and $J = 4$. Note that this exactly describes J th order moving average source modeling.

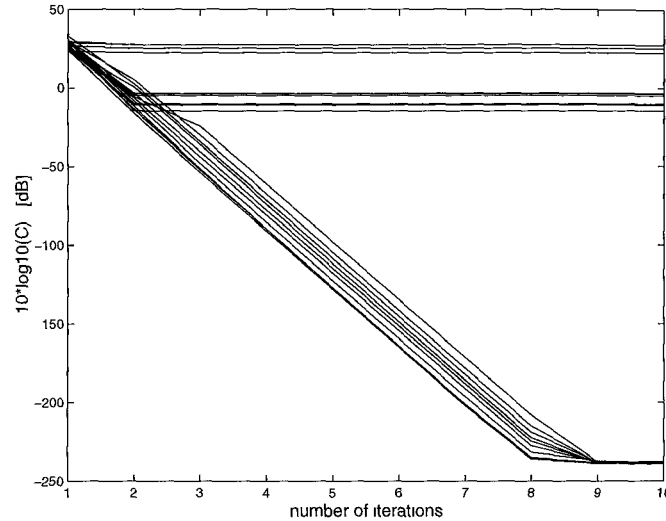


Figure 4.1 Value of objective function on a logarithmic scale versus the number of full iterations.

Initial values for the estimated mixing system were randomly generated by adding Gaussian random variables with standard deviation $\sigma = 0.1$ to the coefficients of the true system. Figure 4.1 shows the convergence behavior for 20 different starting points and the same input statistics. As the examples show, in most cases the value of the objective function decreased to extremely low values. Only three times out of 20 the algorithm got trapped in a local minimum with a relatively high value for C . In all cases where the final value C was below -200 dB the power spectra of the sources were perfectly estimated, and also the mixing system was identified up to permutations and scaling. For cases where the final value was around -14 dB the estimates were close to the true values. For one of these cases, four pairs of the identified source power spectra are depicted in Figure 4.2 together with the true ones. As can be seen, even for these cases the estimated source power spectra are very close to the true ones. Also the cascades of the estimated mixing systems and the

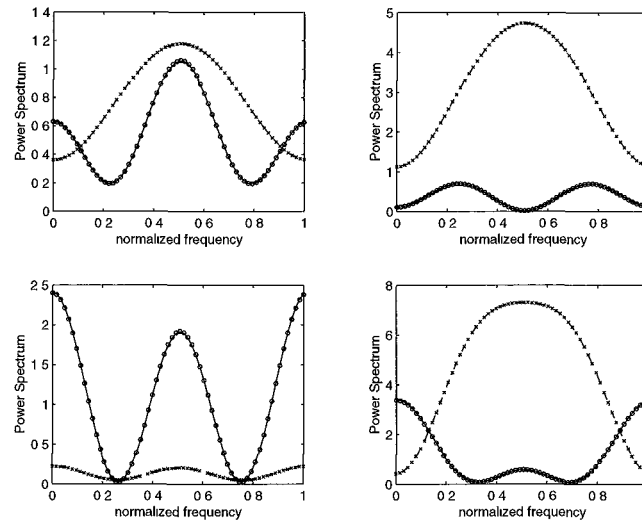


Figure 4.2 Source power spectral densities (psd's) and their estimates Legend: — psd1; - - - psd2; o estimate for psd1, x estimate for psd2.

inverse to the true one were near-perfect, as can be seen in the following example:

$$\mathbf{C}(0) = \begin{bmatrix} 0.0007 & 0.0002 \\ 0.0007 & 0.0002 \end{bmatrix}$$

$$\mathbf{C}(1) = \begin{bmatrix} -1 & 0.0008 \\ 0.0009 & -1 \end{bmatrix} \quad (4.30)$$

$$\mathbf{C}(2) = \begin{bmatrix} 0.0012 & -0.0002 \\ -0.0012 & 0.0002 \end{bmatrix}$$

4.3 Fullband TDBSS Approach

4.3.1 General Overview

Blind Signal Separation (BSS) (K. Pope and R. Bogner, 1996a; K. Pope and R. Bogner, 1996b) has been a topic which attracted many researchers in recent years. With the advent of more powerful processors and the ability to realize more complex algorithms BSS has found useful applications in the areas of audio processing such as speech recognition, audio interfaces, and hands free telephony in reverberant environments. In view of the exponential growth of mobile users in the wireless-communications world together with the limited capacity of resources available for data transmission, modern communication systems increasingly require training-less adaptation, to save on bandwidth capacity or to accommodate unpredictable channel changes. Future systems must utilize spatial diversity multiple access techniques that obtain their channel information exclusively from the received signal. These systems fit the instantaneous and convolutional BSS models. Blind algorithms are useful here as they can be self-recovering and do not require *a priori* knowledge of any training sequence (Feng and Kammeyer, 1998). For example communication systems such as GSM can devote up to 22% of their transmission time to pilot tones which could be otherwise used for data transmission (T. Petermann, D. Boss, and K. Kammeyer, 1999). BSS has also found a fruitful application in multimedia modelling, and recent work on modelling combined text/image data for the purpose of cross-media retrieval has been made using ICA (Larsen et al., 2003).

There is an abundance of various methods used to solve BSS problems, as described in Chapter 2, and these are often application dependent, however; Section 4.3 investigates an algorithm which demonstrates the convolutional mixing model which is relevant to the applications mentioned above and provides a method that avoids the

frequency domain permutation problem. The most prevalent of the aforementioned applications suitable for this particular BSS criterion is in the area of speech processing as it exploits the nonstationarity assumption of the algorithm.

We extend approaches in (M. Joho and K. Rahbar, 2002) to the convolutional mixing cases. In Section 4.3.2 the approaches in (M. Joho and K. Rahbar, 2002) are briefly reviewed. The proposed extended approach for convolutional mixing cases is given in Section 4.3.3. Section 4.3.4 presents the simulation results giving the performance of two local optimization methods: Gradient, and Newton optimization with speech data.

The following notations are used in Section 4.3. We use bold upper and lowercase letters to show matrices and vectors, respectively in the time, frequency and z domains, e.g., $\mathbf{A}(t)$, $\mathbf{A}(\omega)$, $\mathbf{A}(z)$ for matrices and $\mathbf{a}(t)$ for vectors. Matrix and vector transpose, complex conjugation, and Hermitian transpose are denoted by $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H \triangleq ((\cdot)^*)^T$, respectively. $E\{\cdot\}$ means the expectation operation. $\|\cdot\|_F$ is the Frobenius norm of a matrix. \otimes is the Kronecker product and $\text{Trace}(\mathbf{A})$ is the trace of matrix \mathbf{A} . With $\mathbf{a} = \text{diag}(\mathbf{A})$ we obtain a vector whose elements are the diagonal elements of \mathbf{A} and $\text{diag}(\mathbf{a})$ is a square diagonal matrix which contains the elements of \mathbf{a} . $\text{ddiag}(\mathbf{A})$ is a diagonal matrix where its diagonal elements are the same as the diagonal elements of \mathbf{A} and

$$\text{off}(\mathbf{A}) \triangleq \mathbf{A} - \text{ddiag}(\mathbf{A}). \quad (4.31)$$

$\mathbf{1}_{N \times N}$ is an $N \times N$ matrix of ones, $\mathbf{0}_{N \times N}$ is an $N \times N$ matrix of zeros, and \mathbf{I}_N is the $N \times N$ identity matrix. $\text{vec}(\mathbf{A})$ forms a column vector by stacking the columns of \mathbf{A} . The operator $\text{mat}_{N, MQ}(\mathbf{a})$ reshapes a vector \mathbf{a} of length NMQ to an $N \times MQ$ matrix. The matrices \mathbf{P}_{off} , \mathbf{P}_{diag} , and $\mathbf{P}_{\text{vec}}^{(N, L)}$ in Table 4.1 are mainly defined in

accordance with (M. Joho and K. Rahbar, 2002). \mathbf{P}_{off} and \mathbf{P}_{diag} are given by

$$\mathbf{P}_{off} = \text{diag}(\text{vec}(\text{off}(\mathbf{1}_{N \times N}))), \quad (4.32)$$

$$\mathbf{P}_{diag} = \text{diag}(\text{vec}(\mathbf{I}_N)). \quad (4.33)$$

The matrix $\mathbf{P}_{vec}^{(N,L)}$ is the permutation matrix defined by

$$\mathbf{P}_{vec}^{(N,L)} \text{vec}(\mathbf{A}^T) = \text{vec}(\mathbf{A}), \quad (4.34)$$

for $N \times L$ matrices \mathbf{A} . Note that for $N \neq L$ the matrix $\mathbf{P}_{vec}^{(N,L)}$ is, in general, not self-inverse like the one that occurs in (M. Joho and K. Rahbar, 2002).

4.3.2 Review of Instantaneous BSS

This section gives a brief review of the instantaneous BSS algorithm proposed in (M. Joho and K. Rahbar, 2002) which will be extended to our proposed fullband convolutional time domain BSS algorithm given in Section 4.3.3. Assuming that the sources are statistically independent and nonstationary, we observe the measured signals over K different time window frames. In each frame the SOS are considered stationary, while between adjacent frames the SOS are slowly changing. We define the following noise free instantaneous BSS problem. It should be noted that if noise is to be considered, denoising methods that utilize wavelets is a common method that can be incorporated into the BSS process. In the instantaneous mixing cases both the mixing and demixing matrices are constant, that is, $\mathbf{H}(t) = \mathbf{H}$ and $\mathbf{W}(t) = \mathbf{W}$. In this case the reconstructed signal vector can be expressed as

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t). \quad (4.35)$$

The instantaneous correlation matrix of $\hat{\mathbf{s}}(t)$ at time frame k can be obtained as

$$\mathbf{R}_{ss,k} = \mathbf{W}\mathbf{R}_{xx,k}\mathbf{W}^H, \quad (4.36)$$

where $\mathbf{R}_{xx,k}$ is defined as,

$$\mathbf{R}_{xx,k} = E\{\mathbf{x}(k)\mathbf{x}^H(k)\}. \quad (4.37)$$

For a given set of K observed instantaneous correlation matrices, $\{\mathbf{R}_{xx,k}\}_{k=1}^K$, the aim is to find a matrix \mathbf{W} that minimizes the following cost function

$$\mathcal{J}_1 \triangleq \sum_{k=1}^K \beta_k \|\text{off}(\mathbf{W}\mathbf{R}_{xx,k}\mathbf{W}^H)\|_F^2, \quad (4.38)$$

where $\{\beta_k\}$ are positive weighting *normalization* factors such that the cost function is independent of the absolute norms and are given as

$$\sum_{k=1}^K \beta_k \|\mathbf{R}_{xx,k}\|_F^2 = 1. \quad (4.39)$$

Perfect joint diagonalization is possible under the condition that $\{\mathbf{R}_{xx,k}\} = \mathbf{H}\{\mathbf{\Lambda}_{ss,k}\}\mathbf{H}^H$ where $\{\mathbf{\Lambda}_{ss,k}\}$ are diagonal matrices due to the assumption of the mutually independent unknown sources. This means that full diagonalization is possible, and when this is achieved, the cost function given in Equation (4.38) is zero at its global minimum. This constrained non-linear nonconvex multivariate optimization problem can be solved using various techniques including steepest gradient-based descent (SGD), Newton and global optimization routines. However, the performance of the first two techniques depends on the initial guess of the global minimum, which in turn relies heavily on an initialization of the unknown system that is near the global trough. If this is not the case then the solution may be sub-optimal as the algorithm gets trapped in one of the local multi-minima points. Global optimization routines such as those that utilize tunnelling, smoothing, simulated annealing and a combination of first and second order methods allow a more robust convergence of the cost function to the global minimum without additional requirements or *priori* knowledge of source locations with respect to the sensors for good initialization. The area of global optimization with respect to BSS problems remains an open problem as geometric

beamforming and the use of generalized eigenvector decomposition or matrix pencils (C. Chang, Z. Ding, S. Yau, and F. Chan, 2000) are more generally used for good initialization when solving the inherent permutation indeterminacy for convolutive BSS in the frequency domain. The area of global optimization with reference to BSS will be looked at in more detail in Chapter 5.

To prevent a trivial solution where $\mathbf{W} = \mathbf{0}$ would minimize Equation (4.38), some constraints need to be placed on the unknown system \mathbf{W} to prevent this. One possible constraint is that \mathbf{W} is unitary. This can be implemented as a penalty term such as given below

$$\mathcal{J}_2 \triangleq \|\mathbf{W}\mathbf{W}^H - \mathbf{I}\|_F^2, \quad (4.40)$$

or as a hard constraint that is incorporated into the adaptation step in the optimization routine. For problems where the unknown system is constrained to be unitary, Manton (J. Manton, 2002) presented a routine for computing the Newton step on the manifold of unitary matrices referred to as the *complex Stiefel manifold*. For further information on derivation and implementation of this hard constraint refer to (M. Joho and K. Rahbar, 2002) and references therein.

The closed form analytical expressions for first and second order information used for gradient and Hessian expressions in optimization routines and the learning rules for the instantaneous BSS algorithm can be found in (M. Joho and K. Rahbar, 2002). Both the steepest gradient descent (SGD) and Newton methods are implemented following the same frameworks used by (M. Joho and K. Rahbar, 2002).

4.3.3 Proposed Convolutive TDBSS

Most BSS algorithms that assume convolutive mixing reformulate the problem into many problems in the frequency domain using a Fourier transform. By solving many BSS problems in the frequency domain for individual frequency bins one can exploit

the same algorithm derivation as the instantaneous mixing BSS algorithm given in the previous section and also referred to in (M. Joho and K. Rahbar, 2002). However the inherent *frequency permutation problem* remains a problem and will always need to be addressed. The tradeoff is that by formulating algorithms in the frequency domain we can perform less computations and processing time falls but we still must fix the permutations for individual frequency bins so that they are all aligned correctly. This section aims to provide a way to utilize the existing algorithm developed for instantaneous BSS from (M. Joho and K. Rahbar, 2002) and apply it to convolutional mixing but avoid the permutation problem.

We extend the above instantaneous approach to the convolutional case whilst remaining in the time domain. Section 2.4 gives the necessary background to model a typical MIMO convolutionally mixed BSS system. We still assume that the demixing system and reconstructed signals are defined by Equation (2.8), however; we want to maintain similar expressions to those generated in the instantaneous case. It can be shown that Equation (2.8) can be written in the following matrix form

$$\hat{\mathbf{s}}(n) = \mathcal{W}\mathcal{X}(n), \quad (4.41)$$

where $\mathcal{W} \in \mathbb{C}^{N \times QM}$ is given by

$$\mathcal{W} = [\mathbf{W}(0), \mathbf{W}(1), \dots, \mathbf{W}(Q-1)], \quad (4.42)$$

and $\mathcal{X}(n) \in \mathbb{C}^{QM \times 1}$ is defined as

$$\mathcal{X}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-(Q-1)) \end{bmatrix}. \quad (4.43)$$

The output correlation matrix at time frame k can be derived as

$$\mathbf{R}_{ss,k}(0) = \mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(0)\mathcal{W}^T, \quad (4.44)$$

where,

$$\mathbf{R}_{\mathcal{X}\mathcal{X},k}(0) = E\{\mathcal{X}(k)\mathcal{X}^T(k)\} \quad (4.45)$$

Correlation matrices for the recovered sources for all necessary time lags are given by

$$\mathbf{R}_{ss,k}(\tau) = \mathcal{W}E\{\mathcal{X}(k)\mathcal{X}^H(k+\tau)\}\mathcal{W}^H = \mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\mathcal{W}^H. \quad (4.46)$$

Using the joint-diagonalization criterion in (M. Joho and K. Rahbar, 2002) for the instantaneous modelling of the BSS problem, we can formulate a similar expression for convolutional mixing in the time domain. Considering the correlation matrices with all different time lags we should have the following cost function

$$\mathcal{J}_3 \triangleq \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \beta_{k,\tau} \|\text{off}(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\mathcal{W}^H)\|_F^2. \quad (4.47)$$

The only difference between \mathcal{J}_1 and \mathcal{J}_3 is that we now take into account all the different time lags τ for the correlation matrices for each respective time window k where the SOS are changing. Also $\beta_{k,\tau}$ is now defined as

$$\sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \beta_{k,\tau} \|\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\|_F^2 = 1, \quad (4.48)$$

and we note the new structure of \mathcal{W} . In the ideal case where we know the exact demixing system \mathcal{W}_{ideal} , all off-diagonal elements would approximately equal zero and the value of the objective function would reach its global minimum where $\mathcal{J}_3 = 0$. Each value of k represents a different time window frame where the SOS are considered stationary over that particular time frame. In adjacent non-overlapping time frames k and $k+1$, the SOS are changing due to the non-stationarity assumption. As this is a non-linear constrained optimization problem with NQM unknown parameters we can rewrite it as

$$\mathcal{W}_{opt} = \arg \min_{\mathcal{W}} \mathcal{J}_3(\mathcal{W}) \quad (4.49)$$

$$s/t \mathcal{J}_4 = \|\text{ddiag}(\mathcal{W}\mathcal{W}^H - \mathbf{I})\|_F^2 = 0.$$

Due to the structure of the matrices and with the technique of matrix multiplication to perform convolution in the time domain, optimization algorithms similar to those performed in the instantaneous case can be utilized. Notice also that in the instantaneous version the constraint used to prevent the trivial solution, $\mathbf{W} = 0$ was a unitary one. In the convolutional case a different constraint \mathcal{J}_4 is used where the row vectors of \mathcal{W} are normalized to have length one. Again referring to the SGD and Newton algorithms closed form analytical expressions of the gradient and Hessian deduced by Joho and Rahbar (M. Joho and K. Rahbar, 2002) were extended slightly to accommodate the time domain convolutional case of the new algorithm. These expressions are shown in Table 4.1. They are similar to the method proposed in (M. Joho and K. Rahbar, 2002), however, they work for convolutional mixing in the time domain. Full derivations of the gradient and hessian closed form analytical expressions are given in Appendix A.1. For convenience, $\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)$ is denoted as $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau$ for the fullband time domain. With these expressions the SGD and Newton methods are summarized in the Tables 4.2 and 4.3 respectively. Table 4.2 is relatively easy to interpret as it is a simple iterative update or learning rule with a fixed step size. As an alternative to a constant step-size μ the natural gradient method proposed by Amari (S. Amari, 1998) could be used instead of the absolute gradient although faster convergence can be expected from second-order methods. Table 4.3 gives the general Newton update with penalty terms incorporated to ensure that the Hessian of the constraint, denoted as \mathbf{H}_4 , and the gradient of the constraint, denoted as \mathbf{G}_4 , are accounted for in the optimization process. Note the \mathcal{J}_4 defines the constraint given in Equation (4.49) and expresses the unit energy of the rows of \mathcal{W} .

Table 4.1

Closed form analytical expressions for the gradient and Hessian of the cost function and row-normalization constraint.

Cost function - \mathcal{J}_3
$\mathcal{J}_3 = \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \beta_{k,\tau} \ off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H)\ _F^2$
Gradient - \mathbf{G}_3
$\mathbf{G}_3 = 2 \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \beta_{k,\tau} \{off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} + off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau\}$
Hessian - \mathbf{H}_3
$\begin{aligned} \mathbf{H}_3 = 2 \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \beta_{k,\tau} \{ & (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \otimes off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H)) \\ & + (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \otimes off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} \mathcal{W}^H)) \\ & + (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{off}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau *} \otimes \mathbf{I}_N) \\ & + (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau *} \mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{off}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \otimes \mathbf{I}_N) \\ & + (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H \otimes \mathbf{I}_N) \mathbf{P}_{vec}^{(N,N)} \mathbf{P}_{off}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau *} \otimes \mathbf{I}_N) \\ & + (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} \mathcal{W}^H \otimes \mathbf{I}_N) \mathbf{P}_{off} \mathbf{P}_{vec}^{(N,N)} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \otimes \mathbf{I}_N) \} \end{aligned}$
Row-normalized Constraint \mathcal{J}_4
$\mathcal{J}_4 = \ ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I}_N)\ _F^2$
Constraint Gradient \mathbf{G}_4
$\mathbf{G}_4 = 4 ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I}_N) \mathcal{W}$
Constraint Hessian \mathbf{H}_4
$\begin{aligned} \mathbf{H}_4 = 4 (\mathbf{I}_{MQ} \otimes ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I}_N)) \\ & + 4 (\mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{diag}(\mathcal{W}^* \otimes \mathbf{I}_N) \\ & + 2 \mathbf{P}_{vec}^{(N,MQ)} (\mathbf{I}_N \otimes \mathcal{W}^H) \mathbf{P}_{diag}(\mathcal{W}^* \otimes \mathbf{I}_N) \\ & + 2 (\mathcal{W}^H \otimes \mathbf{I}_N) \mathbf{P}_{diag}(\mathbf{I}_N \otimes \mathcal{W}^*) (\mathbf{P}_{vec}^{(N,MQ)})^T \end{aligned}$

Table 4.2

Gradient descent subband BSS algorithm for the joint-diagonalization task with a weighted constraint.

Initialization ($r = 0$) $\cdot \mathcal{W}_0$

For $r = 1, 2, \dots$, till convergence

$$\mathbf{w}_r = \mu \{ \text{vec}(\mathbf{G}_3 + \alpha \mathbf{G}_4) \}$$

$$\Delta \mathcal{W}_r = \text{mat}_{N, MQ}(\mathbf{w}_r)$$

$$\mathcal{W}_{r+1} = \mathcal{W}_r - \Delta \mathcal{W}_r$$

Table 4.3

Newton-type subband BSS algorithm for the joint-diagonalization task with a weighted constraint.

Initialization ($r = 0$) $\cdot \mathcal{W}_0$

For $r = 1, 2, \dots$, till convergence

$$\mathbf{w}_r = \mu(\mathbf{H}_3 + \alpha \mathbf{H}_4)^{-1} \text{vec}(\mathbf{G}_3 + \alpha \mathbf{G}_4)$$

$$\Delta \mathcal{W}_r = \text{mat}_{N, MQ}(\mathbf{w}_r)$$

$$\mathcal{W}_{r+1} = \mathcal{W}_r - \Delta \mathcal{W}_r$$

4.3.4 Simulation Results

To demonstrate the performance of the extended convolutional BSS algorithm in the time domain we firstly investigated the instantaneous BSS algorithm using a variety of optimization techniques. A set of $K = 15$ real diagonal square matrices $\{\Lambda_k\}$ were randomly chosen representing the unknown source input uncorrelated matrices. The diagonal assumption is crucial to all BSS problems as it reflects that the sources are mutually independent allowing separation. Following the assumption that the unknown separating system \mathbf{W} is unitary, preventing the trivial solution, the observed correlation matrices can be constructed where $\{\mathbf{R}_{xx,k}\} = \{\mathbf{H}\Lambda_k\mathbf{H}^H\}$ and

\mathbf{H} is chosen as a two by two unitary mixing matrix,

$$\mathbf{H} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \quad (4.50)$$

This first simulation compares the different optimization methods used in (M. Joho and K. Rahbar, 2002) for the instantaneous case. Equation (4.38) forms the objective to be optimized while Equation (4.40) forms the constraint preventing a trivial solution of the unknown separating system \mathbf{W} . Figure 4.3 shows the comparison of the convergence rates of each optimization algorithm for ten independent runs with ten distinct sets of correlation matrices. It is evident that with the second order information available, convergence of the Newton algorithm is much quicker. The optimization for this particular instantaneous BSS problem where the system is assumed to be unitary is performed on the Stiefel manifold. The step size $\mu = 0.2$ was used and the various slopes of the different convergence curves of the gradient method depends entirely on the ten different sets of randomly generated diagonal input matrices.

With the SGD and Newton methods, convergence to the global minimum depends entirely on a good initial starting point \mathbf{W}_0 . The starting point selected in this simulation was

$$\mathbf{W}_0 = \begin{bmatrix} \cos(1) & -\sin(1) \\ \sin(1) & \cos(1) \end{bmatrix}. \quad (4.51)$$

Although techniques like geometric beamforming (L. Parra and C. Alvino, 2002; S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001) are good procedures to allow feasible starting points for optimization, they require additional information and assumptions on the problem space. To investigate the performance of the extended instantaneous BSS algorithm to the convolutional case in the time domain the SGD and Newton algorithm implementations in (M. Joho and K. Rahbar, 2002) were altered to the learning rules given in Tables 4.2 and 4.3 respectively. As the constraint no

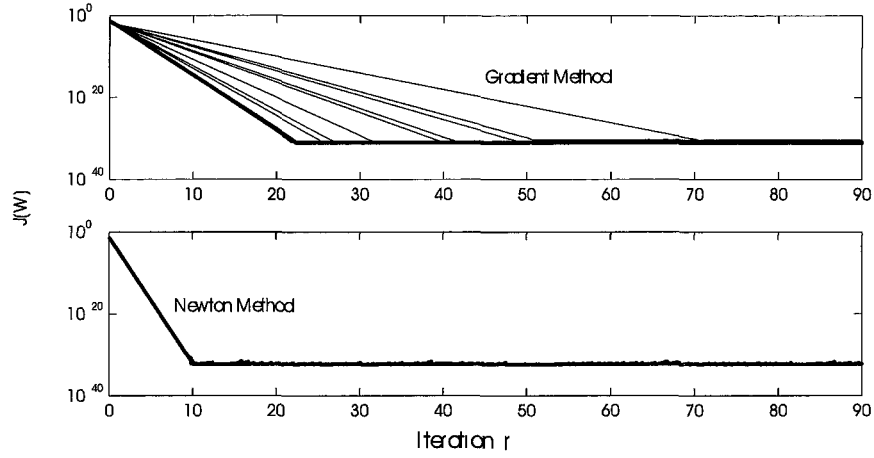


Figure 4.3 Convergence of differing optimization methods for instantaneous BSS.

longer requires the unknown system \mathcal{W} to be unitary, the constraint was changed to that given in Equation (4.49). The technique of weighted penalty functions was used to ensure the constraints preventing the trivial solution were met. No longer performing the optimization on the Stiefel manifold as in (M. Joho and K. Rahbar, 2002) the SGD and Newton algorithms were changed to better reflect the row normalization constraint for the convolutive case. Using the causal z -transform

$$\mathbf{H}_{ij}(z) = \sum_{n=0}^{\infty} \mathbf{h}_{ij}(n) z^{-n}, \quad \forall i, \forall j, \quad (4.52)$$

a first-order, $P = 2$, two-input-two-output (TITO) two tap FIR known mixing system was chosen and is given below in the z domain as

$$\mathbf{H}(z) = \begin{bmatrix} 1 + z^{-1} & -1 + z^{-1} \\ -1 + z^{-1} & 1 + z^{-1} \end{bmatrix} \quad (4.53)$$

The corresponding known demixing system which would separate mixed signals which are produced by convolving the source signals with the TITO mixing system

$\mathbf{H}(z)$ given above is

$$\mathbf{W}_{ideal}(z) = \frac{1}{4} \begin{bmatrix} 1 + z^{-1} & 1 - z^{-1} \\ 1 - z^{-1} & 1 + z^{-1} \end{bmatrix}. \quad (4.54)$$

This is the exact known inverse multiple-input-multiple-output (MIMO) FIR system of the same order. The convolution of these two systems in cascade would ensure the global system $\mathbf{G}(z) = \mathbf{W}_{ideal}(z)\mathbf{H}(z)$ would be a delayed version of the identity, i.e. $z^{-1}\mathbf{I}$. Using matrix multiplication to perform convolution in the time domain, Equation (4.42) can be used to represent the equivalent structure of Equation (4.54),

$$\frac{1}{4}\mathcal{W}_{ideal} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}. \quad (4.55)$$

Through empirical analysis we set the parameters $\mu = 0.6$ and $\alpha = 0.2$ and solve the constrained optimization problem given in Equation (4.49) using the SGD and Newton methods. A set of $K = 15$ real diagonal square uncorrelated matrices for the unknown source input signals were randomly generated. Using convolution in the time domain a corresponding set of correlation matrices $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau$ for each respective time instant $k = 1, \dots, 15$ at multiple time lags τ were generated for the observed signals. Each optimization algorithm was run ten independent times and convergence graphs were observed and are shown in Figure 4.4. The various slopes of the different convergence curves of the gradient method depends entirely on the ten different sets of randomly generated diagonal input matrices. Poor initial values for the unknown system lead to convergence to local minima as opposed to the desired global minimum. The initialization of the SGD and Newton algorithms plays an important role in the convergence to either a local or global minimum. Initial values for the estimated demixing system \mathcal{W} were generated using a perturbed version of the true demixing system. This was done by adding Gaussian random variables with standard deviation $\sigma = 0.1$ to the coefficients of the true system. In most cases this

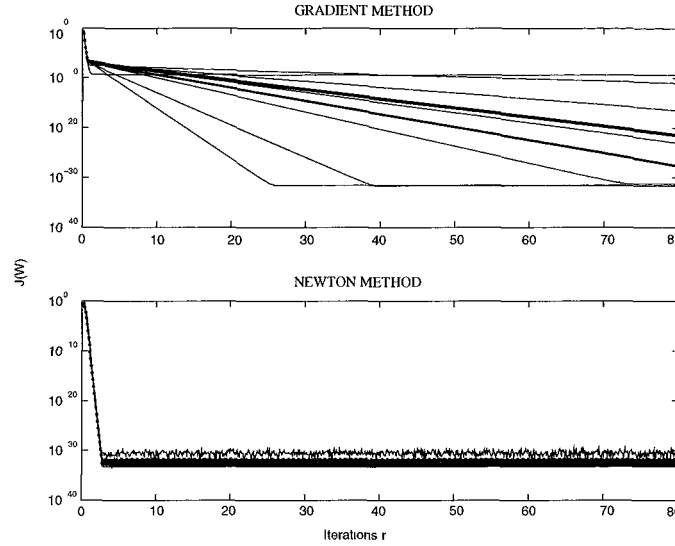


Figure 4.4 Convergence of gradient descent and Newton algorithms for a first order TITO FIR demixing system over 10 trials

initial value is set by exploiting some sort of *a priori* knowledge about the problem such as geometrical layout of sources in relation to the sensors.

After convergence of the objective function to an order of magnitude approximately equal to 10^{-34} the unknown demixing FIR filter system \mathcal{W} , in cascade with the known mixing system $\mathbf{H}(z)$, resulted in a global system which was equivalent to a scaled and permuted version of the true global system $z^{-1}\mathbf{I}$ as can be seen by the

following example,

$$\begin{aligned}\mathbf{G}(0) &= \begin{bmatrix} -0.17 & 0.17 \\ 0.19 & -0.19 \end{bmatrix} \times 10^{-14}, \\ \mathbf{G}(1) &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \\ \mathbf{G}(2) &= \begin{bmatrix} -0.23 & -0.23 \\ -0.14 & -0.14 \end{bmatrix} \times 10^{-14}\end{aligned}\tag{4.56}$$

A first order system has been identified up to an arbitrary global permutation and scaling factor. The TITO system identified above using the optimization algorithms has only 8 unknown variables to identify. We now examine a MIMO FIR mixing system with a higher dimension. Again we have chosen an analytical MIMO multivariate system whose exact FIR inverse is known. The 3rd order mixing system is given below in the z domain

$$\mathbf{H}_{11}(z) = -4 - 4z^{-1} + z^{-2} + z^{-3},\tag{4.57}$$

$$\mathbf{H}_{12}(z) = -7 - 7z^{-1} + z^{-3},\tag{4.58}$$

$$\mathbf{H}_{21}(z) = 7 - 7z^{-1} + z^{-3},\tag{4.59}$$

$$\mathbf{H}_{22}(z) = 9 - 9z^{-1} - z^{-2} + z^{-3}\tag{4.60}$$

The corresponding known inverse FIR system of the same order is given below also in the z domain as

$$\mathbf{W}_{11}^{ideal}(z) = \frac{1}{13}\mathbf{H}_{22}(z),\tag{4.61}$$

$$\mathbf{W}_{12}^{ideal}(z) = -\frac{1}{13}\mathbf{H}_{12}(z),\tag{4.62}$$

$$\mathbf{W}_{21}^{ideal}(z) = -\frac{1}{13}\mathbf{H}_{21}(z),\tag{4.63}$$

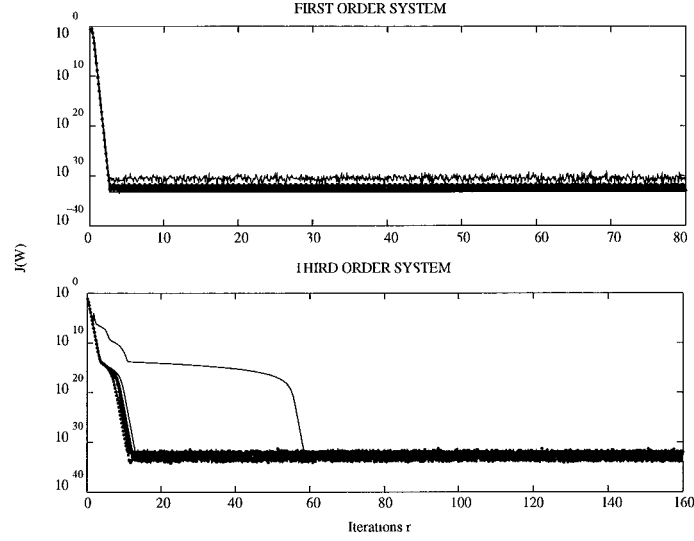


Figure 4.5 Convergence of Newton algorithms for first and third order TITO FIR demixing systems over 10 trials.

$$\mathbf{W}_{22}^{ideal}(z) = \frac{1}{13} \mathbf{H}_{11}(z). \quad (4.64)$$

The convolution of the mixing and demixing MIMO FIR systems given in Equations (4.57-4.64) gives the identity matrix \mathbf{I} exactly. A comparison of the convergence behaviour for the more efficient Newton method is given in Figure 4.5 using the same methods described for the first order systems above, keeping the learning factor and weighting terms the same. We see from the figure that with twice as many unknown variables to solve for the demixing system, the third order unknown system takes longer to converge by roughly a factor of two. Both systems converge to their global minimums due to good initialization at approximately 10^{-34} . For the third order system, one trial produced an outlying convergence curve that takes more iterations r than the other trials. This is dependent on the randomly generated set of diagonal correlation matrices $\{\mathbf{R}_{ss,k}\}$ where $k = 1, 2, \dots, 15$ for each trial.

To test the performance of the algorithm on real speech data two independent segments of speech were used as input signals to the MIMO FIR mixing system given in

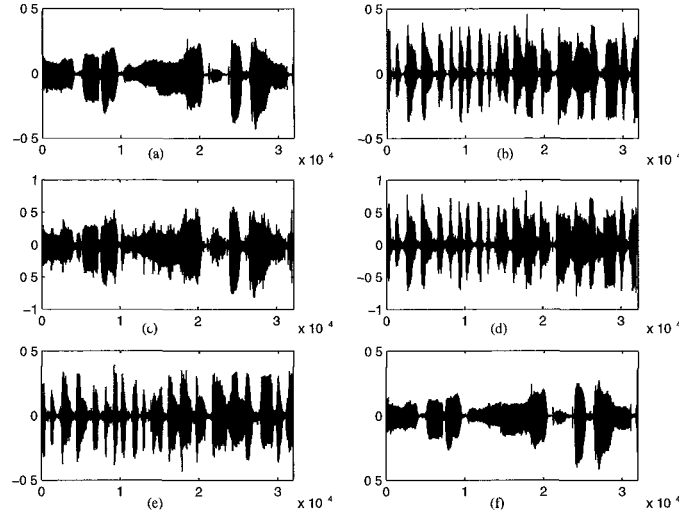


Figure 4.6 (a) and (b) are the two original signals, (c) and (d) are the convolutively mixed signals, (e) and (f) are the permuted separated results.

Equation (4.53). These signals were both 4 seconds long and sampled at 8kHz. The signals were convolutively mixed with the synthetic mixing system to obtain 2 mixed signals. With the assumption that speech is quasi-stationary over a period of approximately 20ms, the observed mixed signals were buffered and segmented into 401 frames each having 160 samples in length. The nonstationarity assumption assumes that the SOS in each frame does not change. The correlation matrices $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau$ can be found via Equations (4.45,4.46) for $K = 401$ frames of the two mixed signals. This allows the method of joint diagonalization by minimizing the off-diagonal elements of the correlation matrices of the recovered signals at each respective time lag τ as defined in Equations (4.47,4.49). Figure 4.6 shows the input, mixed and recovered speech signals. A good qualitative recovery is confirmed by subjective listening to the recovered audio signals and inspection of graphs (e) and (f) in Figure 4.6.

4.4 Conclusions

This chapter has proposed two new BSS algorithms in a convolutive mixing environment. Firstly a new method for the blind estimation of convolutive mixing systems in the presence of colored sources has been presented with an ACDC ALS algorithm which avoids the local frequency permutation problem. The results show that the algorithm estimates both the source spectra and the mixing system with little variance if it converges to a low value of the objective function. Further work will be directed toward automatic initialization of the algorithm and optimization for recorded data such as speech and audio. Also a new method for convolutive BSS in the time domain extending upon an existing instantaneous BSS framework has been presented. This method avoids the inherent permutation problem when dealing with solving the convolutive BSS problem in the frequency domain. Optimization algorithms including SGD and Newton methods have been compared. Future work will be directed at implementing the simulations where the orders of the mixing and demixing MIMO FIR systems are higher.

Chapter 5

Uninitialized BSS with Global Optimization

5.1 Introduction

In this chapter we discuss the efficiency and implementation details of applying the branch-and-bound global optimization algorithm Dividing Rectangles (DIRECT) (D. R. Jones, DIRECT, 2001; M. Bjorkman and K. Holmström, 1999; D. R. Jones, C. D. Perttunen, and B. E. Stuckman, 1992) with clustering (K. Holmström, and M. Edvall, 2004) to solve an uninitialized convolutive blind signal separation (BSS) problem driven by nonstationary sources in the time domain (TD) for the multiple-input-multiple-output (MIMO) system case as proposed in Chapter 4. Current methods for solving such multi-modal BSS problems involve the use of local optimization techniques as described in Chapter 2. The main motivation for using global optimization approaches such as the one described in this chapter for BSS problems is that we cannot always assume *a priori* knowledge. Current methods, as described in Chapter 2, obtain information using additional assumptions on the problem to ensure good initialization in the region of attraction, or basin, of the global minimum so as to prevent convergence to a local minimum corresponding to a sub-optimal solution.

By using global optimization we eliminate the need for additional assumptions for good initialization of the unknown demixing system and solve the problem where the global minimum corresponds to the optimal demixing system. The BSS algorithm implemented in the time domain in Chapter 4, avoids the frequency permutation problem and has less computational overhead than the typically used frequency domain method for a small to medium number of dimensions. The use of global optimization using the DIRECT algorithm is compared to a local optimization method. The number of function evaluations, iterations till convergence and quality of separation for all methods is examined. Computational complexity of both a frequency domain method with initialization and sorting of permutations, and the proposed time domain algorithm using global optimization is also examined. The benefits of using global optimization when compared to local optimization techniques in BSS problems as far as complexity goes is justified for small to medium sized demixing systems but the primary motivation is where initialization information is not readily available to the problem. We firstly give a brief mathematical description of the generic DIRECT optimization algorithm. Then we see how the algorithm is adapted to accommodate the solving of our proposed multi-modal uninitialized BSS problem. Finally some simulations comparing the proposed optimization method and time domain algorithm from Chapter 4 to a typical frequency domain approach with local optimization using additional information in the problem space are given.

In all problems, optimization of the criteria is essential and there are many algorithms which use additional assumptions or *a priori* knowledge for good initial starting points for local optimization methods. Also the traditional approach to convolutive BSS problems is to transform the problem into the frequency domain and solve many instantaneous problems and then sort the separated outputs for each respective frequency bin to ensure correct alignment across the entire frequency spectrum for cor-

rect reconstruction of the unknown input signals (L. Parra and C. Alvino, 2002). For our particular problem we assume a multiple-input-multiple-output (MIMO) convolutive mixing system of FIR filters. A backward model is used to generate the separated nonstationary source signals. Source signals are assumed to be nonstationary on the justification that most real-world signals are inherently nonstationary or cyclostationary. The time domain algorithm formulated in Chapter 4 is used in conjunction with the proposed global optimization framework of this chapter. This algorithm exploits the nonstationarity of the input signals and uses a method of joint-diagonalization to minimize the off-diagonal elements of the correlation matrices of the reconstructed signals over K different time window frames. In each of these frames the SOS are stationary, while between adjacent frames, k , and $k + 1$, the SOS are slowly changing. The algorithm avoids the permutation problem and is a more effective approach to solving the demixing system where the number of dimensions is relatively small to medium scale. In this chapter the input signals and mixing/demixing system can be real or complex-valued and the MIMO channel impulse responses are time-invariant over a finite interval. When looking at local methods of optimization, closed form analytical expressions for the gradient and Hessian of the objective function, as well as the constraint Jacobian, and Hessian of the second term of the Lagrange function, are taken from Table 4.1 with proofs given in Appendix A.1. The demixing MIMO system that performs deconvolution on the observed signals is found when we solve for the global minimizer of the constrained multivariate BSS optimization problem.

The main applications of interest in which this problem can be applied is in audio signal processing and communications systems. Multiple acoustic nonstationary signals recorded simultaneously in a reverberant environment by multiple microphones have a mixing system that can be modelled by the convolutive MIMO system of FIR

filters. For highly reverberant environments where the reverberation time is high, estimation of demixing systems requires estimation of demixing FIR filters of several thousand taps when sampling at 8kHz, and incorporation of some sort of transform to the frequency domain (L. Parra and C. Spence, 2000) or the subband domain is necessary. Array beamforming is generally used for successful initialization for local optimization methods.

Wireless communications systems where signal separation plays an important role is in applications using wireless multi-user mobile communication systems where shared use of the same frequency band by mobile users is made (N. Sellami, M. Siala, and I. Fijalkow, 2004; Feng and Kammeyer, 1998). For MIMO multipath fading channels that are flat (frequency non-selective), we can model the multipath channels using the instantaneous mixing model for BSS. However if the channel is assumed to be non-flat (frequency selective) then we can assume a MIMO system of FIR filters can be used to model the channel (J. Choi, 2004). In applications such as these, multiple antennas, utilizing spatial diversity, can transmit cyclostationary digital signals such as symbol streams, or space-time block codes, through a multipath propagation MIMO channel, impinging on a uniform linear antenna array. These received signals must be equalized so as to recover the transmitted signals. Typically some training sequence known by the transmitter and receiver, such as pilot and data symbols, is sent periodically to identify the unknown mixing channel. This overhead reduces available bandwidth that could be otherwise utilized for the user's data. If the geometry of the antenna manifold is known we obtain good initialization for local optimization. In both of the above cases where we cannot assume additional knowledge, global optimization methods for BSS prove useful.

There are numerous global optimization methods available throughout literature to solve multimodal nonconvex nonlinearly constrained optimization problems without

any priori information for an initial starting point in the basin of attraction of the global minimum(s). Methods that utilize the Lipschitz constant, simulated annealing, genetic evolution, multistart searches, random uniform distributed search, clustering and P-algorithms are ubiquitous in the field of global optimization and a good review for various methods is given in (R. Horst, and P. M. Pardalos, 1995). However few approaches have been considered when dealing with BSS problems. The main categories of global optimization include methods with guarantee of some tolerance threshold and Bayesian methods. In particular, the branch-and-bound method derived from the areas of interval analysis and combinatorial optimization is implemented to solve the time domain BSS objective function. A guarantee is provided to find the global minimizer x^* with a desired accuracy after a predictable number of steps. The method does not require the calculation of a derivative or Hessian, although efficiency is improved when expressions for these, such as in Table 4.1, are available; this will be investigated in more detail in Section 5.2.3. The problem space is bounded and is split recursively by branching into smaller and smaller sections based on some preferential selection system and is guaranteed to converge within tolerance thresholds if the objective is continuous in the basin of attraction of the global minimizer.

5.2 DiRect Algorithm

The DIRECT algorithm is one particular type of branch-and-bound method which uses an adaptive combination of global and local optimization methods. A complete mathematical description of the generic DiRect algorithm from (M. Björkman and K. Holmström, 1999) is provided in Table 5.1 for completeness. The explanation for parameters used in Table 5.1 are provided in Table 5.2. A simple summary of the main steps of the algorithm is presented in Table 5.3 with more elaboration on these

steps given in the following sections when using the algorithm with reference to the BSS problem.

The basic steps of the algorithm given in Table 5.3 will be briefly summarized but for a more detailed description of the algorithm, refer to (M. Bjorkman and K. Holmström, 1999). The original version of the DIRECT algorithm solved problems of the form given by Equation (5.1) without the nonlinear constraint $c(x)$. For our constrained optimization problem, a newer version of DIRECT that accounts for such constraints $c(x)$ is implemented using the *glcFast* function from (Holmström, 2002) based on ideas from (D. R. Jones, DIRECT, 2001) and is given by

$$\begin{aligned}
 \min \quad & f(x) \\
 \text{s.t.} \quad & x_L \leq x \leq x_U \\
 & b_L \leq Ax \leq b_U \\
 & c_L \leq c(x) \leq c_U
 \end{aligned} \tag{5.1}$$

where $x, x_L, x_U \in \mathbb{R}^n$, $c(x), c_L, c_U \in \mathbb{R}^{m_1}$, $A \in \mathbb{R}^{m_2 \times n}$ and $b_L, b_U \in \mathbb{R}^{m_2}$, where m_1 and m_2 are the number of nonlinear and linear constraints respectively. It should be noted that in our implementation of BSS to acoustic applications we use a cosine-

modulated FIR filter bank that decomposes the real observed signals into real subband component signals. Within each subband the DIRECT algorithm can optimize with respect to a real demixing system. To apply the DIRECT algorithm when the unknown demixing system is complex, some minor alterations to the implementation steps of the algorithm need to be made but this is left for future work. For the complex demixing system case, this also doubles the amount of variables to solve for. This may occur when using a DFT modulated FIR filter bank for decomposition of real observed signals into complex subband components, or when the observed signals are digital modulation signals such as QAM, QPSK as found in communication applications.

The search space is limited to an n -dimensional unit hypercube. The algorithm proceeds by partitioning the hypercube into smaller hyper-rectangles, by trisection, each having a sampling point at its center, i.e. a point where the objective function is to be evaluated. Figure 5.1 shows an example of partitioning on a hypothetical problem

when the number of dimensions is $n = 2$. Figure 5.2 shows a hypothetical example of the division process for $n = 3$. Each iteration t begins by selecting one or more of the hyper-rectangles for further search. The function is evaluated at the new points and then the new potential hyper-rectangles are chosen and subdivided with the new points for evaluation being the center of the new hyper-rectangles. The selection criteria from (D. R. Jones, C. D. Perttunen, and B. E. Stuckman, 1992) is as follows,

Selection Criterion. *Select a rectangle for further search if there exists an f^* satisfying*

$$(f_r - f^*)/d_r \leq (f_s - f^*)/d_s, \forall s \neq r \quad (5.2)$$

$$f^* \leq f_{mn} - \varepsilon, \quad (5.3)$$

where $\varepsilon > 0$ is the desired accuracy. Note that f_r and f_s denote the function value at the center of hyper-rectangles r and s respectively, where hyper-rectangle s is the previous good hyper-rectangle and hyper-rectangle r is the current one. d_r and d_s denote the distance from the center of the hyper-rectangle to its vertices and ε is the tolerance error to which we find our global minimum up to. For our particular problem we know that the global optimum occurs when the value of our objective function in Equation (4.47) has $\mathcal{J}_3 = 0$. This additional information allows us to assume we know the function value of the global minimum f^* . Finding the set of potentially optimal hyper-rectangles S is equivalent to finding the lower-right convex hull of a particular set of points on a plane. This becomes a subproblem of the algorithm. Convex hulls are briefly described in the Section 5.2.1, but for further detailed reference on this part of the algorithm refer to (D. R. Jones, DIRECT, 2001; M. Björkman and K. Holmström, 1999; F. Preparata, and M. Shamos, 1985). After finding all the optimal sub-hyper-rectangles we repeat the process until our updated value f_{mn} is within the tolerance threshold or we reach a pre-determined number of iterations or number of function evaluations. The corresponding approximate global

minimizer x_{min} will correspond to the value of the final best sample point or center point of the most optimal sub-hyper-rectangle. Once this point is identified we can conclude that x_{min} is in the basin of attraction of the global minimizer and so a local search can be performed for further refinement if desired. This particular method of optimization is adaptive and combines both a global and local search method eliminating both the cumbersome slow convergence of a pure global method such as a random search and the unreliability of a pure local method that may become trapped in a local multiminima point. The global method intelligently continually identifies basins of convergence to a better minimum.

If we compare the general criterion in Equation (5.1) to our proposed time domain BSS problem in Equation (4.49) we see that we must place bounds on the unknown demixing system \mathcal{W} . Also it is seen that there is no linear equality constraints and only one nonlinear constraint to prevent the trivial solution.

5.2.1 Convex Hulls

In general, if S is a set of points in a plane, then the convex hull of S is the smallest convex polygon that contains all points of S . When selecting hyper-rectangles if we only used a global search then we would select one of the biggest hyper-rectangles in each iteration i.e. point B in Figure 5.3. By doing this, the hyper-rectangles would become smaller at about the same rate and sampled points would represent a uniform grid in the n -dimensional normalized hypercube search space. In (D. R. Jones, DIRECT, 2001), to avoid this, a balance between global and local behavior is introduced by selecting not one but many hyper-rectangles. A relative size measuring the distance from the center of each hyper-rectangle to each of the vertices is given. It is hard to visualize this concept in space-time for more than 3 or 4 dimensions. Figure 5.3 from (D. R. Jones, DIRECT, 2001) shows a plot of points with each point corresponding to one of the hyper-rectangles in the partition for one iteration. The abscissa represents the size/distance for each hyper-rectangle from smallest to largest. The ordinate is the function value corresponding to the sampled middle points of the hyper-rectangles. DIRECT selects the points on the lower-right convex hull indicated in the figure corresponding to the potentially optimal hyper-rectangles selected for further sub-division.

5.2.2 Clustering

When midpoints are selected from optimal hyper-rectangles, these points appear in clusters. The function *glcCluster* from (Holmström, 2002) incorporates the benefits of using a clustering algorithm with the DIRECT algorithm to eliminate redundancy of identifying points belonging to the same basin of attraction. i.e. points that will converge to the same minimum. For details on the clustering algorithm used in our implementation refer to (K. Holmström, and M. Edvall, 2004). Basically the cluster-

ing algorithm is integrated with DIRECT as follows:

1. Identify feasible points using DIRECT algorithm
2. Apply a clustering algorithm on all sampled points by DIRECT to identify a set of clusters. The point with the lowest function value in each cluster is selected
3. Perform a local search on each of the best cluster points as initial starting values
4. If the best point found from local searches is better than best point found in

the initial global search, we perform DIRECT again with an initial value corresponding to the best point found in this step

5. If the DIRECT algorithm with initial value is better than in step 4, we repeat again until we are satisfied.

5.2.3 Sequential Quadratic Programming

In this section we discuss the method of local optimization that is incorporated into the DIRECT algorithm. As shown in Tables 4.2 and 4.3, the local optimization methods employed were the steepest gradient descent method and the Newton method. Both of these methods used the closed form analytical expressions of Table 4.1. The use of penalty terms for both optimization methods was used to ensure the constraint \mathcal{J}_4 given in Equation (4.49) was met. Here the penalty term α was chosen heuristically and the step size parameter μ was kept constant during each iteration. Naturally the Newton method converged much quicker than the typically used gradient descent method but required an initial value which was close to the solution to prevent ill-

convergence to local multim minima. The local optimization solver used in conjunction with *glcCluster* as part of the DIRECT algorithm is the solver *snopt* from (Holmström, 2002) which stands for Sparse Nonlinear Optimization. This routine uses a large-scale sequential quadratic programming (SQP) sub solver and assumes sparse linear or nonlinear constraints. The problem specified in Equations (4.49) and (5.1) fits this model. SNOPT uses the SQP algorithm to obtain search directions from a sequence of quadratic programming subproblems. Each QP subproblem minimizes a quadratic model of a certain Lagrangian function subject to a linearization of the nonlinear constraints. An augmented Lagrangian merit function is reduced along each search direction to ensure convergence from any starting point (Gill, 2002). Known expressions of first and second order derivatives of the objective function and constraints are provided in Table 4.1. If these are not specified then they are calculated using finite differences or automatic differentiation. As in Equation (5.1) there are bounds placed on the variables and there is a nonlinear equality constraint $c(x)$, which in our problem is \mathcal{J}_4 . The merit function for step-length, (i.e. μ), control is an augmented Lagrangian. A brief summary of the SQP method is given but for a more detailed mathematical explanation refer to (Gill, 2002) and references therein.

The basic process of SNOPT involves major and minor iterations. Each major iteration r is a QP subproblem where a search direction is chosen towards the next iterate $r + 1$. The constraints of the subproblem are formed from the linear constraints, in our problem there are none, and the nonlinear constraint linearization. The QP subproblem is to minimize a quadratic approximation to a modified Lagrangian function subject to some QP constraints. Solving the QP subproblem at each iteration is itself an iterative procedure with the minor iterations of a SQP method being each iteration of the QP method. After a QP subproblem has been solved for one major iterate, new estimates of the nonlinear programming (NP) solution are computed using a

line search procedure employing a cubic interpolation method, on the augmented Lagrangian merit function. The line search determines the adaptive step size μ_r for that particular iterate r where $0 < \mu_r < 1$ such that the new point gives a sufficient decrease in the merit function. Other local optimization solvers were explored such as *conSolve* which implements the SQP algorithm by Schittkowski with Augmented Lagrangian merit function described in (P. E. Gill, W. Murray, and M. A. Saunders, 1997), and *nlpSolve* which implements the Fletcher SQP algorithm described in (R. Fletcher, and S. Leyffer, 1997). Both of these methods though are sub-optimal compared to the *snopt* solver described above.

5.3 Simulation Results

The local methods used in Chapter 4 were the steepest gradient descent and the Newton methods. Both methods utilized exact closed form analytical expressions for the gradients and Hessian of the objective function \mathcal{J}_3 and constraint function \mathcal{J}_4 , given in Equation (4.49). Here we will denote these to be $\mathbf{G}_3, \mathbf{G}_4, \mathbf{H}_3$, and \mathbf{H}_4 respectively. In Chapter 4 we used the learning rules for gradient and Newton methods given in Table 4.2 and 4.3. Also the constraint was incorporated as a penalty term. The penalty coefficient α was determined heuristically. Also the step size parameter μ was kept constant and also determined heuristically. In this chapter to perform the initialized local optimization the solver *snopt* from (Holmström, 2002) is used. This routine uses a sequential quadratic programming (SQP) algorithm. The algorithm obtains search directions from a sequence of quadratic programming subproblems. At each major iteration, an approximation is made of the Hessian of the Lagrangian function using a quasi-Newton updating method. A search direction is found and then a line search procedure using a cubic interpolation method to find the adaptive step size for each search direction is used. Unlike the methods used in chapter 4, the

search step μ is not constant and changes with each iteration. Also the penalty term α is found exactly using a Lagrangian multiplier. This proves to be more optimal than the methods used in chapter 4. For a more detailed description of the SQP algorithm refer to (P. E. Gill, W. Murray, and M. A. Saunders, 1997). In our simulation we use a first order FIR filter MIMO mixing system also used in Chapter 4 with $n = 8$ dimensions. We define it as

$$\mathbf{H}(z) = \begin{bmatrix} 1 + z^{-1}, & -1 + z^{-1} \\ -1 + z^{-1}, & 1 + z^{-1} \end{bmatrix}. \quad (5.4)$$

As in Chapter 4, the corresponding known demixing system which would separate mixed signals which are produced by convolving the source signals $s(t)$ with the TITO FIR mixing system $\mathbf{H}(z)$ is given as

$$\mathbf{W}_{ideal}(z) = \frac{1}{4} \begin{bmatrix} 1 + z^{-1}, & 1 - z^{-1} \\ 1 - z^{-1}, & 1 + z^{-1} \end{bmatrix} \quad (5.5)$$

This is the exact known inverse TITO FIR system of the same order. The convolution of these two systems in cascade would ensure the global system, $\mathbf{G}(z) = \mathbf{W}_{ideal}(z)\mathbf{H}(z)$, would be a delayed version of the identity, i.e. $z^{-1}\mathbf{I}$. To test the convergence of the algorithms we generate a set of $K = 15$ real diagonal square uncorrelated matrices $\{\mathbf{R}_{ss,k}\}$. From this we convolve with the mixing system to obtain $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^T$. For the global optimization approach we do not initialize, however for the local optimization approach we use a perturbed version of the true demixing system as the good starting point to ensure convergence to the correct solution. Normally this is done using some geometric beamforming method. Figure 5.4 shows convergence information about the local methods with good and bad initialization as well as the global method with no initialization. The local method with good initialization converges in 26 iterations while the global method converges in 41 iterations. We can see that even with a small variation in the starting point for a local method of optimization on a nonconvex problem with many multiminima, we can become trapped

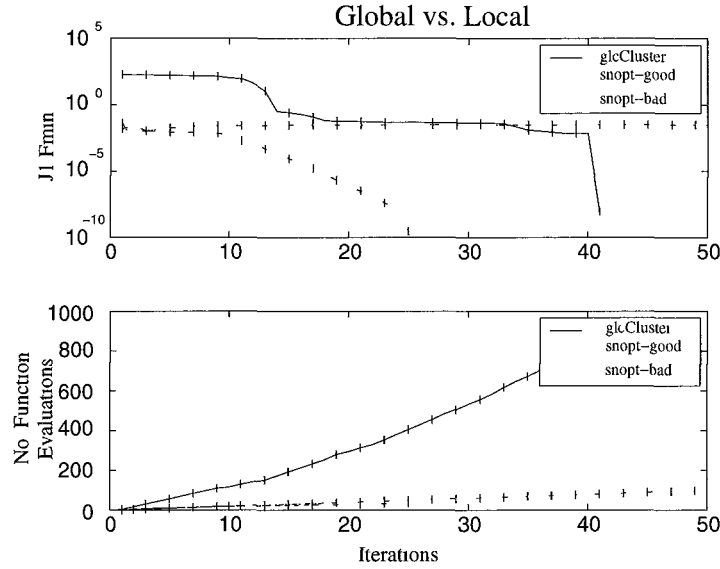


Figure 5.4 Comparison of global and local optimization routines, glcCluster and snopt

in a local minima and we do not obtain the correct solution. For problems where we cannot assume additional knowledge of the problem we see the advantages of using a global approach. The resultant global system for the bad initialized local method and the global method is shown in Equations (5.6) and (5.7) respectively. The global system for the good initialized local method is very similar to Equation (5.7).

$$\mathbf{G}(0) = \begin{bmatrix} 0.733 & -0.733 \\ -0.735 & 0.735 \end{bmatrix},$$

$$\mathbf{G}(1) = \begin{bmatrix} 0.009 & 1.364 \\ 0.005 & 1.364 \end{bmatrix}, \quad (5.6)$$

$$\mathbf{G}(2) = \begin{bmatrix} -0.728 & -0.728 \\ 0.727 & 0.727 \end{bmatrix}.$$

$$\begin{aligned}
\mathbf{G}(0) &= \begin{bmatrix} -0.663 & 0.663 \\ 0.907 & -0.907 \end{bmatrix} \times 10^{-9}, \\
\mathbf{G}(1) &= \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}, \\
\mathbf{G}(2) &= \begin{bmatrix} -0.798 & -0.798 \\ -0.398 & -0.398 \end{bmatrix} \times 10^{-8}.
\end{aligned} \tag{5.7}$$

5.4 Benefits for Small to Medium Scale Systems

When you keep a cost function for convolutive BSS in the time domain, then you can elegantly avoid a permutation problem arising in the algorithm. When dealing primarily with acoustical signals, the demixing filters of interest for dereverberation are a few thousand taps long and large scale optimization methods generally have a high computational cost due to computing analytical expressions of Hessian and Jacobian matrices of objective and constraint functions making these methods unfeasible. However if the applications that a particular convolutive BSS algorithm is being applied to fall into the communications area, then unknown demixing filters are generally much shorter and the benefits of not having to solve the permutation problem are evident with less computations as there is no transform function or permutation sorting/correction. Recently in (M. Joho, 2004), a time-domain algorithm was developed for convolutive BSS that avoids the inherent frequency permutation problem. This algorithm derives the objective function over all cross-correlations over all time-lags, and the update-equation, entirely in the time domain and carries out fast convolution in the frequency domain. Still, the algorithm performs transforms of the

data from the time to frequency domain and back again a few times which increases computational overhead. In Figure 5.5, a comparison of computations for a typical initialized frequency domain method using geometric beamformers in (L. Parra and C. Spence, 2000; L. Parra and C. Alvino, 2002), with the uninitialized convolutive BSS time domain method using *glcCluster* is given. It illustrates that for demixing FIR filter MIMO systems where the number of taps is relatively small, the computational complexity, measured by the *flops* function in Matlab v5.3, is less for the *glcCluster* method when the number of dimensions is less than approximately 36, corresponding to a FIR filter length of 9 taps for a TITO system. To benefit from this time-domain convolutive uninitialized BSS problem using the *glcCluster* and *snopt* solvers we require that the unknown demixing system(s) be a small to medium number of dimensions. To achieve this for acoustic applications where the number of variables in the fullband is a few thousand, it is necessary to perform subband decomposition on the observed fullband mixed signals using a uniform FIR filter bank with oversampling to obtain subband components that have fewer variables to

solve for within each respective subband. Decoupled subbands will still result with the permutation problem which must be solved, however as realized in (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001; M. Ikram, and D. Morgan, 2000; R. Mukai, S. Araki, H. Sawada, and S. Makino, 2004; S. Araki, S. Makino, R. Aichner, T. Nishikawa and H. Sarawatari, 2003) motivation for adopting subband based BSS is due to the upper bound placed on separation performance for other typical convolutive BSS methods.

5.5 Conclusions

In this chapter we have presented a method to solve convolutive BSS problems driven by nonstationary signals where no additional information is available to provide a good initial starting point in the region of the global minimizer corresponding to the true demixing system. A global optimization algorithm DIRECT is used on a time domain BSS model that avoids the frequency permutation problem that other convolutive BSS models have to solve. The benefits of using this approach to solve for demixing systems that have a low to medium number of dimensions are seen by comparing our approach to a frequency domain BSS model that solves the permutation problem using a projection operation and that uses additional assumptions on *a priori* knowledge along with geometric beamforming to obtain a good initializer for the multivariate nonconvex problem.

Chapter 6

Subband BSS Model

6.1 Introduction

In this chapter we solve blind signal separation (BSS) of nonstationary convolutively mixed source signals in an acoustic environment using a subband domain model for the general multiple-input-multiple-output (MIMO) system case. The motivation for subband based BSS is due to the upper bound placed on separation performance due to equivalence between adaptive beamformers (ABF) and BSS when using the typical frequency domain BSS approach for highly reverberant environments as shown in (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001; M. Ikram, and D. Morgan, 2000; R. Mukai, S. Araki, H. Sawada, and S. Makino, 2004) and discussed in Chapter 2. A subband approach also solves the permutation indeterminacy more effectively than a frequency domain BSS approach when long reverberation times are considered by using FIR filters in each subband as shown in (S. Araki, S. Makino, R. Aichner, T. Nishikawa and H. Sarawatari, 2003). Oversampled \tilde{M} channel FIR filter banks using both DFT modulation and cosine modulation designs are used in conjunction with the proposed time domain blind source separation (BSS) algorithm given in Chapter 4. This BSS algorithm has been shown to blindly separate the fullband versions of non-stationary convolutively mixed sources in the time domain.

However further savings on convergence and computational complexity can be made by using subband decomposition on the mixed signals before implementation of the time domain BSS algorithm in each subband. An extended lapped transform (ELT) prototype is modulated using a cosine-modulated (CM) FIR filter bank and then with a DFT modulated FIR filter bank. Both of these designs are compared to the typical frequency domain BSS approach to solving these convolutive non-stationary BSS problems such as in (L. Parra and C. Spence, 2000). A new adjacent subband coupling metric is used with a dyadic sorting routine to detect and fix permutations amongst subbands before the synthesizing stage of the filter bank. The introduction of global optimization using the Dividing Rectangles (DIRECT) algorithm (D. R. Jones, DIRECT, 2001; M. Björkman and K. Holmström, 1999; D. R. Jones, C. D. Perttunen, and B. E. Stuckman, 1992) with clustering (K. Holmström, and M. Edvall, 2004), given in Chapter 5 for solving uninitialized BSS problems in individual subbands, is integrated into the subband domain where no additional assumptions are required for estimating good initializers based on geometric beamforming, directivity patterns, or known direction-of-arrival (DOA) angles, as is the case in (L. Parra and C. Alvino, 2002; R. Mukai, H. Sawada, S. Araki, and S. Makino, 2004; W. Wang, J. Chambers, and S. Sanei, 2004; H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004a). ICA'99 benchmark data for both synthetic and real test cases are used for simulations and evaluation indicators such as the signal-to-interference (SIR) ratio is used when comparing the proposed model to a typical frequency domain BSS method.

For BSS problems that have convolutive mixing systems that model real environments using MIMO FIR filters, the number of unknown variables that must be estimated is in the order of several thousand. Traditionally these convolutive BSS models are solved by transforming to the frequency domain such as in (L. Parra

and C. Spence, 2000). As an alternative, we are motivated to investigate different methods of separation by including a subband preprocessor before implementing the time domain BSS algorithm given in chapter 4. To reduce the convergence time for solving the total number of unknown parameters in the fullband model, subband decomposition is performed as a preprocessor to the time domain BSS algorithm thus solving in the subband domain as opposed to the fullband. This is implemented using oversampled uniform filter bank models satisfying perfect reconstruction (PR), including DFT and CM FIR filter banks such as in (J. Kliewer, and A. Mertins, 1998; Koilpillai and Vaidyanathan, 1992). These two models are then compared to the traditional frequency domain BSS method given in (L. Parra and C. Spence, 2000) and the separation performance for each model is measured.

In Section 6.2 a typical frequency domain method to solving convolutive BSS problems is given. References to the limitations of such methods are also made. Section 6.3 defines the oversampled uniform filter bank models including the CM and DFT modulated filter banks based on an ELT prototype. Section 6.4 integrates the filter banks with the fullband time domain convolutive BSS algorithm to allow subband BSS with global optimization. Section 6.5 gives a comparative analysis of the subband based BSS models with a traditional frequency domain method with focus on the separation performance using the SIR BSS metric. The real mixing response of a typical office room is measured and identified. This identified system is mixed synthetically with segments of real speech signals taken from the TIMIT corpus of speech to produce some mixed signals. These mixed signals are used as input to each of the three convolutive BSS models and initialization of the unknown demixing system to be identified is a perturbed version of the known demixing system. Section 6.6 proposes a new subband coupling metric which solves the local permutation problem before the synthesis stage of the filter bank. A dyadic sorting routine for aligning

adjacent subbands across the entire spectrum is provided in Section 6.7. Simulation results are presented in Section 6.8 and the conclusions are given in Section 6.9.

6.2 Frequency Domain Methods and Limitations

Typically for the convolutive mixing BSS model described in Section 2.4, the time-domain convolutive problem is transformed into independent, multiple short-term or instantaneous mixing, BSS problems in the frequency domain via the T -point discrete Fourier transform (DFT). For the frame $[\mathbf{x}(t), \dots, \mathbf{x}(t + T)]$ this is given as $\mathbf{x}(\omega, t) = \sum_{\tau=0}^{T-1} e^{-i2\pi\omega\tau/T} \mathbf{x}(t + \tau)$. In a compact matrix-vector notation the time-frequency domain relationships are shown as

$$\mathbf{x}(t) = \mathbf{H}(t) * \mathbf{s}(t) \quad (6.1)$$

which in the frequency domain becomes,

$$\mathbf{x}(\omega) = \mathbf{H}(\omega) \mathbf{s}(\omega) \quad (6.2)$$

The reconstructed signals in the frequency domain are defined as

$$\hat{\mathbf{s}}(\omega) = \mathbf{W}(\omega) \mathbf{x}(\omega), \quad (6.3)$$

which corresponds to the original source signals, in the frequency domain, up to an arbitrary permutation and scaling factor; that is

$$\hat{\mathbf{s}}(\omega) = \mathbf{W}(\omega) \mathbf{H}(\omega) \mathbf{s}(\omega) = \mathbf{\Pi}(\omega) \mathbf{D}(\omega) \mathbf{s}(\omega) \quad (6.4)$$

In this case, $\mathbf{W}(\omega) \in \mathbb{C}^{N \times M}$ and $\mathbf{H}(\omega) \in \mathbb{C}^{M \times N}$. $\mathbf{D}(\omega) \in \mathbb{C}^{N \times N}$ is an arbitrary frequency dependent diagonal scaling matrix. The permutation matrix $\mathbf{\Pi}(\omega) \in \mathbb{R}^{N \times N}$ is frequency dependent and introduces frequency-dependent permutation errors in the output frequency response. In order to avoid the inherent frequency permutation

problem it is desirable to either make $\mathbf{\Pi}$ independent of frequency or derive a criteria to ensure correct permutation alignment of all separated frequency bins.

A typical convolutive BSS criterion in the frequency domain is to exploit the nonstationary assumption of the source signals and use joint diagonalization on the SOS of the observed signals to minimize the off-diagonal elements of the output cross-power spectra $\mathbf{R}_{ss,k}(\omega)$ using the most recent observed cross-power spectra $\mathbf{R}_{xx,k}(\omega)$ over all frequency bins. The relationship between observed and output cross-power spectra is as follows. The statistically independent source cross-power spectra is given as

$$\mathbf{R}_{ss,k}(\omega) = E\{\mathbf{s}_k(\omega)\mathbf{s}_k(\omega)^H\}, \quad (6.5)$$

where due to independence, ideally the set of $\{\mathbf{R}_{ss,k}(\omega)\} \forall k, \omega$ are diagonal matrices. The observed and output cross-power spectral matrices are given as

$$\mathbf{R}_{xx,k}(\omega) = E\{\mathbf{x}_k(\omega)\mathbf{x}_k(\omega)^H\}, \quad (6.6)$$

and

$$\mathbf{R}_{\hat{s}s,k}(\omega) = \mathbf{W}(\omega)\mathbf{R}_{xx,k}(\omega)\mathbf{W}(\omega)^H, \quad (6.7)$$

respectively. This leads to a criterion for simultaneous diagonalization of $\mathbf{R}_{ss,k}(\omega) \forall k, \omega$.

A possible objective function for the frequency domain approach is given as,

$$\mathcal{J}_5 \triangleq \sum_{\omega=1}^T \sum_{k=1}^K \alpha_{k,\omega} \|off(\mathbf{W}(\omega)\mathbf{R}_{xx,k}(\omega)\mathbf{W}^H(\omega))\|_F^2, \quad (6.8)$$

where

$$\alpha_{k,\omega} = \left(\sum_{k=1}^K \|\mathbf{R}_{xx,k}(\omega)\|_F^2 \right)^{-1}. \quad (6.9)$$

The criterion then can be defined as

$$\mathbf{W}_{opt} = \arg \min_{\mathbf{W}(\omega)} \mathcal{J}_5(\mathbf{W}(\omega)) \quad (6.10)$$

$$s/t \mathcal{J}_6 = \|ddiag(\mathbf{W}(\omega)\mathbf{W}^H(\omega) - \mathbf{I}_N)\|_F^2 = 0,$$

where the constraint \mathcal{J}_6 is used to prevent the trivial solution. This general joint diagonalization model can be found in (L. Parra and C. Alvino, 2002; L. Parra and C. Spence, 2000; M. Joho and K. Rahbar, 2002) with a different constraint to prevent the trivial solution. Other approaches include the use of an alternating least squares (ALS) optimization method which can be found in (K. Rahbar, J. Reilly and J. Manton, 2004) and (A. Yeredor, 2002). However, to prevent the frequency permutation problem additional constraints based on *a priori* knowledge on the spatial geometry of the array manifold, or source/sensor location using blind beamforming are usually introduced for good initialization of the unknown demixing system in each frequency bin, i.e. $\mathbf{W}(\omega)$.

In (L. Parra and C. Spence, 2000), a noise free criterion very similar to that given in Equation (6.10) is given below as

$$E(\omega, k) = \mathbf{W}(\omega) \bar{R}_x(\omega, k) \mathbf{W}^H(\omega) - \Lambda_s(\omega, k) \quad (6.11)$$

$$\begin{aligned} \hat{\mathbf{W}}, \hat{\Lambda}_s = \arg \min_{\mathbf{W}, \Lambda_s,} & \sum_{\omega=1}^T \sum_{k=1}^K \|E(\omega, k)\|^2 \\ & \mathbf{W}(\tau) = 0, \tau > Q \ll T, \\ & W_i(\omega) = 1 \end{aligned} \quad (6.12)$$

For this convolutive BSS frequency domain algorithm, only consistent permutations for all frequencies will correctly reconstruct the sources and the constraint on filter size Q versus the frequency resolution $1/T$ links the otherwise independent frequencies and solves the problem. The filter constraint is achieved via a projection operator that zeros the appropriate delays for every demixing channel. In (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001) and (M. Ikram, and D. Morgan, 2000) the limitations of the frequency domain method are considered when mixing is conducted in reverberant rooms where the microphones are not positioned close to the sources resulting in long demixing FIR filters. It is shown that such a constraint is

not effective for long reverberant environments and that the separation performance is highly limited where such long impulse responses are evident. Other methods found in (L. Parra and C. Alvino, 2002; H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004a) and (H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004b) used to solve the permutation problem utilize geometric constraints or initializers based on positions of sources and sensors to mitigate the problem. Beam patterns or knowledge of the direction of arrival (DOA) of the target source impinging on the sensor array manifold is utilized however when this information is not available other methods used to detect and correct permutation inconsistency between adjacent frequency bins proves useful.

6.3 Subband Model

To utilize BSS in the subband domain we must perform subband decomposition using some type of uniform or non-uniform FIR filter bank. As Figure 6.1 shows, we have chosen an oversampled uniform \tilde{M} channel modulated FIR filter bank in direct form based on the modulation of an ELT prototype function $h(n)$ from (Malvar, 1992) given by

$$h(n) = -\frac{1}{2\sqrt{2}} + \frac{1}{2} \cos \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2\tilde{M}} \right]. \quad (6.13)$$

Two designs of modulated FIR filter banks that exhibit perfect reconstruction (PR) are investigated for subband BSS and these are cosine modulated (CM) filter banks and discrete Fourier transform (DFT) modulated filter banks. We investigate both designs but we are more interested in the former design as it uses real subband components where we do not need to modify the DIRECT algorithm to accommodate complex variables and in our acoustic application we are observing real signals. Without performing any subband processing on the decomposed mixed signals $\mathbf{x}_{\{1,2,\dots,M\}}^{sub}(p, k)$, where k is the time frame index and $p = 0, 1, \dots, \tilde{M} - 1$ is the sub-

band index, we obtain PR. Performing subband BSS using the algorithm referred to in Chapter 4, aliasing is introduced and so we must oversample by the factor $\frac{\tilde{M}}{R}$ to minimise this. Direct form versions of the filter banks are used here for simplicity however equivalent polyphase structures for both types filter banks can improve efficiency (J. Klierer, and A. Mertins, 1998; Koilpillai and Vaidyanathan, 1992). Also further information on implementing efficient fast ELT FIR filter banks with overlapping analysis filters using DCT IV and advantages of using ELT prototype functions can be found in (Malvar, 1992).

6.3.1 Cosine Modulated FB

A filter bank is said to be cosine modulated if all analysis and synthesis filters are generated by cosine modulation of one or two prototype filters. The prototype low-pass filter has a cutoff of $\pm\pi/2\tilde{M}$ for \tilde{M} filters. Individual analysis and synthesis filters have real coefficients and are of equal length. The impulse responses of the synthesis FIR filters are defined as

$$f_p(n) = h(n) \sqrt{\frac{2}{\tilde{M}}} \cos \left[\left(n + \frac{\tilde{M} + 1}{2} \right) \left(p + \frac{1}{2} \right) \frac{\pi}{\tilde{M}} \right], \quad (6.14)$$

and the analysis filters are related as

$$f_p(n) = h_p(L - 1 - n), \quad (6.15)$$

where $p = 0, 1, \dots, \tilde{M} - 1$, and $n = 0, 1, \dots, L - 1$. For the ELT prototype defined in Equation (6.13), $L = 4\tilde{M}$. Due to the oversampling factor $\frac{\tilde{M}}{R}$ to obtain PR of the filter bank a scalar of $\sqrt{\frac{R}{\tilde{M}}}$ must be multiplied with each $f_p(n)$. Figure 6.2 shows the impulse responses for the analysis filters for the first two subbands as well as the magnitude frequency responses of the first three subbands for an ELT CM FIR FB when $\tilde{M} = 8$.

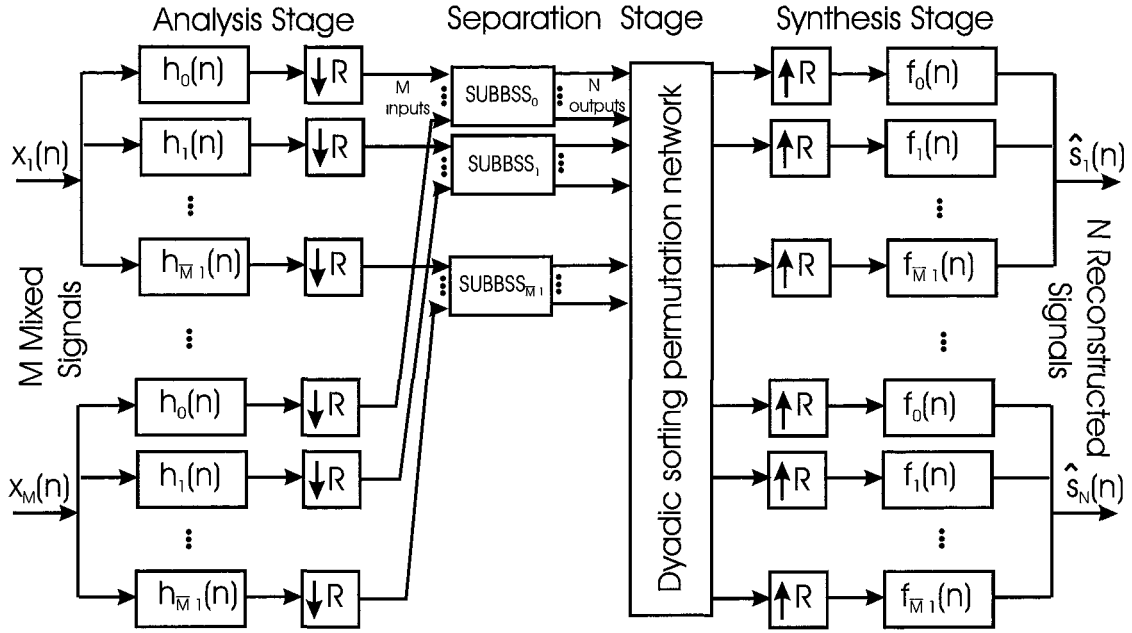


Figure 6.1 General subband MIMO BSS model with oversampling factor $\frac{\tilde{M}}{R}$

6.3.2 DFT Modulated FB

This filter bank uses exponential modulation. The individual analysis and synthesis filters have complex coefficients in the DFT filter design. The prototype lowpass filter has a cutoff of $\pm\pi/\tilde{M}$ for \tilde{M} filters. Note that we consider a $2\tilde{M}$ -band DFT and \tilde{M} band cosine-modulated filter bank so that the subbands are of equal spectral width in both filter bank designs. The impulse response of the synthesis FIR filter is defined as

$$f_p(n) = h(n)e^{j(\frac{2\pi}{\tilde{M}})p(n-\frac{L-1}{2})} \quad (6.16)$$

and the analysis filters are related as

$$f_p(n) = h_p(n), \quad (6.17)$$

where $p = 0, 1, \dots, \tilde{M} - 1$, and $n = 0, 1, \dots, L - 1$ with $L = 2\tilde{M}$. The scalar factor $\sqrt{\frac{2R}{\tilde{M}^2}}$ again must be added due to the oversampling factor.

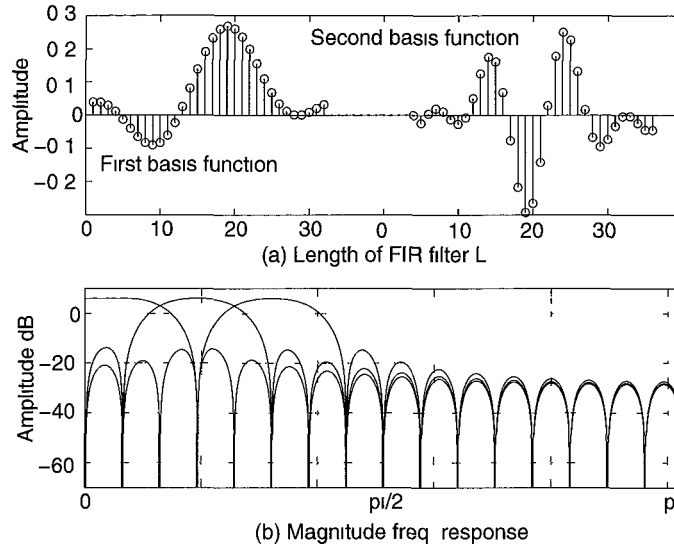


Figure 6.2 (a) Shows the impulse responses for the analysis filters for the first two subbands and (b) shows the magnitude frequency responses of the first three subbands for an ELT CM FIR FB when $\tilde{M} = 8$.

6.4 Integration of TDBSS into Subband Model

There are three stages to the subband model as shown in Figure 6.1. Firstly we decompose the M fullband mixed signals $\mathbf{x}(t)$ into \tilde{M} subbands via the analysis stage of the filter bank to obtain the subband signals $\mathbf{x}_{\{1,2,\dots,M\}}^{sub}(p, k)$ where k is the time frame index and $p = 0, 1, \dots, \tilde{M} - 1$ is the subband index. With the cosine modulated design, the fullband mixed signals are convolved with the respective impulse responses of the analysis filters defined in Equation (6.15) and then oversampled by the factor $\frac{\tilde{M}}{R}$ while convolution with the impulse responses defined in Equation (6.17) provides the DFT modulated result for each subband also after subsampling by R . In Chapter 4 we solve a problem in the fullband domain where there exist ' NMQ ' free parameters. Shorter FIR filters of length Q_p can be solved for each subband using DIRECT which effectively reduces the overall convergence time of the algorithm to find the unknown demixing system. Note that each subband BSS problem

is a MIMO problem where there are M input signals from each respective subband of the mixed signals and N separated output signals for each respective subband. In the second stage, integration of the fullband time domain BSS algorithm given in Chapter 4 is simply made by substituting the subband versions of the mixed signals $\mathbf{x}_{\{1,2,\dots,M\}}^{sub}(p, k)$ and the unknown demixing system $\mathcal{W}^{sub}(p)$, for the fullband versions of the mixed signals $\mathbf{x}(t)$ and the unknown demixing system \mathcal{W} , and solve p separation problems where $p = 0, 1, \dots, \tilde{M} - 1$. For simplicity, $\mathbf{x}_{\{1,2,\dots,M\}}^{sub}(p, k)$ is denoted as $\mathbf{x}_{p,k}$, $\mathcal{W}^{sub}(p)$ is denoted as \mathcal{W}_p , and $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^p(\tau)$ is denoted as $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau,p}$. Substituting \mathcal{W}_p , Q_p and $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau,p}$ for \mathcal{W} , Q and $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau}$ in all expressions in Table 4.1 respectively, will provide the subband expressions for $\mathcal{J}_{3,p}$, $\mathbf{G}_{3,p}$, $\mathbf{H}_{3,p}$, $\mathcal{J}_{4,p}$, $\mathbf{G}_{4,p}$ and $\mathbf{H}_{4,p}$. It should be noted that the value of Q_p will be determined by the decided value of Q in the fullband domain, the number of subbands \tilde{M} , the length of the analysis FIR filters L and the oversampling ratio $\frac{\tilde{M}}{R}$. A more detailed explanation of choosing the subband filter length Q_p can be found in (J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, 2002). Before the final stage we correct any permutation ambiguities and ensure consistent scaling between adjacent separated subbands for each signal. The final stage of the model is the synthesis stage and involves upsampling the separated subband signals $\hat{\mathbf{s}}_{\{1,2,\dots,N\}}^{sub}(p, k)$ by R and convolving this result with the respective impulse responses of the synthesis filters defined in Equation (6.14) and (6.16). This will provide the N fullband separated signals $\hat{\mathbf{s}}(t)$. For each subband BSS problem the global optimization method DIRECT is used. To overcome the local subband permutation problem when geometric information is unavailable, a dyadic sorting routine used to align all subbands to the same permutation is used and described in Section 6.7.

6.5 Comparing Subband and Frequency Domain BSS Models

In this section we report the results of separation of two mixed signals in a realistic environment such as an office room given in Figure 3.4 using the three different models described in Section 6.3. As described in Chapter 3, we identify the MIMO convolutive mixing impulse responses $\mathbf{H}_{known}(\tau)$ coupling two loudspeakers and two microphones in a reverberant environment. The technique used to obtain the corresponding known demixing impulse responses for separation $\mathbf{W}_{known}(\tau)$, or equivalently \mathcal{W}_{known} , is described in Chapter 3. Using $8kHz$ as the sampling rate, \mathcal{W}_{known} has a FIR filter length of $Q = 2048$ for a response time of $T_R = 250ms$. The two input signals $\mathbf{s}(t)$ are speech segments taken from the TIMIT corpus of speech. These signals are convolutively mixed with $\mathbf{H}_{known}(\tau)$ and provide the mixed signals $\mathbf{x}(t)$ which are observed by the two cardioid microphones that have an inter-element spacing of $38cm$. These mixed signals are used for each particular algorithm.

For the cosine modulated FIR filter bank model we decompose the unknown fullband demixing system \mathcal{W} into $\tilde{M} = 256$ subbands with a subsampling factor of $R = 64$. This will mean that instead of trying to solve the nonlinearly constrained optimization problem given in Equation (4.49) for $'NMQ' = 8196$ unknown variables we have in each subband only 192 variables to solve for. Similarly for the DFT modulated model we decompose the unknown fullband demixing system into $\tilde{M} = 512$ subbands with a subsampling factor of $R = 128$. This ensures that the spectral width of the analysis and synthesis filters in the DFT modulated case is the same as that for the cosine modulated case. For this simulation the Newton method time domain BSS algorithm for convolutive mixtures given in Table 4.3 was used to solve for each subband unknown demixing system. To achieve this the fullband mixed signals $\mathbf{x}(t)$

were passed through the analysis stage shown in Figure 6.1 for each subband model to obtain the $\mathbf{x}_{\{0,1,\dots,M\}}^{sub}(p, m)$ subband signals respectively. Initial values of each subband unknown demixing system were set to a perturbed version of the known demixing subband system $\mathcal{W}_{known}^{sub}(p, m)$. For explanation on how this is achieved refer to Section 6.8.2. In most cases information on the demixing system is unknown and geometric beamforming (S. Araki, S. Makino, R. Aichner, T. Nishikawa and H. Sarawatari, 2003) is used to provide initialization information for the optimization process however in this case we are comparing the separation performance of the algorithms and initialization is the same for all three algorithms. In Section 6.8 we examine the integration of the global optimization algorithm presented in Chapter 5. The weighting factor for the penalty term for the constraint in the Newton update from Table 4.3 is set to $\alpha = 0.2$ and the learning coefficient is $\mu = 0.8$. The number of time frames over which joint diagonalization is performed is $K = 128$ which corresponds to a non-stationary time period for speech of 20 – 30 ms. The typical BSS frequency domain approach also sets the required variables of the algorithm to be $Q = 2048, T = 4096, K = 128$ and an initial value for the unknown system in each frequency bin T that is derived by simply taking a T -point Fourier transform of the perturbed known fullband demixing system \mathcal{W}_{known} . In order to evaluate the performance of the proposed BSS methods we used the signal to interference ratio $SIR_i = SIR_{O_i} - SIR_{I_i}$, discussed in Chapter 2 and again defined in Section 6.8. SIR means the ratio of a target-originated signal to a jammer-originated signal (L. Parra and C. Spence, 2000). For the subband BSS models the fullband converged solutions for \mathcal{W} after the synthesis stage are then converted to the frequency domain via a T -point DFT to allow comparison with the BSS frequency domain approach using the SIR metric. Figure 6.3 shows the performance comparison of the two proposed methods with the typical frequency domain method. After each iteration through the algorithms we measure the SIR in decibels for each method. We only

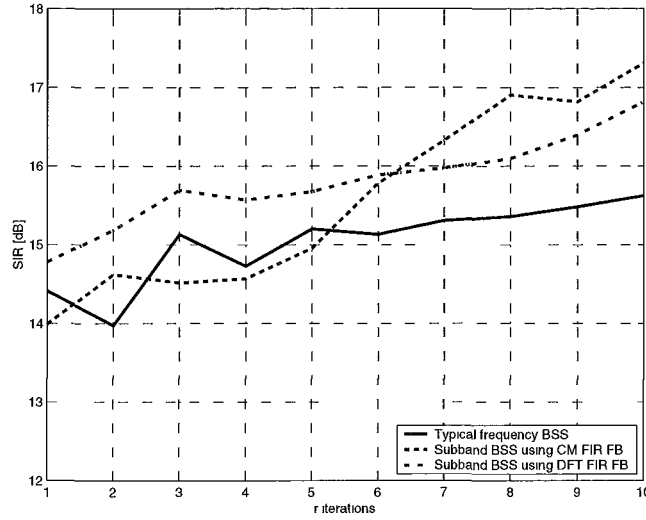


Figure 6.3 Separation performance using three different BSS techniques for two TIMIT speech segments recorded with two cardioid microphones in a reverberant office environment.

look at the first 10 iterations for the three methods. Initially we see that the subband based BSS that uses the DFT FIR filterbank has the highest SIR at 14.85 dB. After the 6th iteration the subband based BSS algorithm using the ELT prototype with CM is better with a higher SIR than the other two methods.

6.6 Subband Coupling Metric

As shown with the frequency domain approach, solving the convolutive BSS problem results in a permutation indeterminacy. With the limitations on separation performance for the frequency domain approach, a subband based approach for convolutive BSS is applied instead. This however will still result with a permutation ambiguity for each separated subband. The subband permutation ambiguity can be defined as

$$\hat{\mathbf{s}}_{\{1,2,\dots,N\}}^{sub}(p, k) = \mathbf{\Pi}_p \mathbf{s}_{\{1,2,\dots,N\}}^{sub}(p, k), \quad (6.18)$$

where $\mathbf{\Pi}_p$ is the subband dependent permutation matrix for the p^{th} subband where $p = 0, 1, \dots, \tilde{M} - 1$. There are two steps to correcting this permutation problem.

Firstly a criterion for detecting a different permutation between consecutive pairs of adjacent subbands is required. A typical approach for resolving permutations between adjacent separated frequency bins in the frequency domain approach is to exploit the cross-frequency correlation or interactions between the separated cross frequency spectra for adjacent frequency bins when the sources are speech signals. In (J. Anemüller, 2001) and (N. Murata, S. Ikeda, and A. Ziehe, 2001), this method was employed exploiting the amplitude modulation correlation across adjacent frequency bands due to the spectrum modulation of speech. A similar metric is derived with respect to separated adjacent subbands for the TITO system case. The derivation of incorrect permutation detection and correction for the general MIMO system will be saved as a future exercise. After correctly identifying a permutation ambiguity between two adjacent subbands after separation, a correction routine must be carried out. The detection and correction steps are described as follows.

After successfully identifying the demixing system of FIR filters \mathcal{W}_p for all subbands $p = 0, 1, \dots, \tilde{M} - 1$, we obtain the separated subband components for each signal. Each solution achieved via global optimization is decoupled and thus the local permutation indeterminacy results between arbitrary separated subbands. We may utilize information from the assumption that our unknown input signals are speech. Over a time frame of 20 ms, speech is considered stationary. Observing the respective decomposed separated signals for the same time frame of speech over all subbands in the ideal case, where there exists no local permutation indeterminacies, there exists a higher correlation between the complex envelopes of adjacent subbands from the same separated signal than there does between the complex envelopes of adjacent subbands from the different signals. Using this fact and assuming the local scaling indeterminacy between separated subbands is solved, we can employ the normalized

cross-correlation metric defined below as

$$\rho_{mn}(p_i, p_j) = \frac{Env\{\hat{s}_m^{sub}(p_i, k)\} * Env\{\hat{s}_n^{sub}(p_j, k)\}}{\sqrt{Env\{\hat{s}_m^{sub}(p_i, k)\}^2} \sqrt{Env\{\hat{s}_n^{sub}(p_j, k)\}^2}} \quad (6.19)$$

where $\rho_{mn}(p_i, p_j)$ represents the cross correlation of the complex envelopes of adjacent subbands p_i and p_j of the k^{th} stationary time frame of speech between the separated subband signals \hat{s}_m^{sub} and \hat{s}_n^{sub} . We define the function to obtain the complex envelope of the respective separated subband signal to be

$$Env\{\hat{s}^{sub}(p, k)\} = |\hat{s}^{sub}(p, k) + j\mathcal{H}\{\hat{s}^{sub}(p, k)\}| * h_0(n), \quad (6.20)$$

where $\mathcal{H}\{\cdot\}$ is the Hilbert function and a low pass filter, h_0 defined in Equation (6.15), is used to obtain the envelope of the subband signal. If the adjacent subbands p_i and p_j have the same permutation then it is expected that

$$\frac{\rho_{11}(p_i, p_j) + \rho_{22}(p_i, p_j)}{\rho_{12}(p_i, p_j) + \rho_{21}(p_i, p_j)} > 1. \quad (6.21)$$

If this is not the case, then we conclude that there is a incorrect permutation at one of the adjacent subbands and must permute one of the subbands to meet the above condition. This correlation ratio detection test for a TITO system can be applied over all consecutive adjacent separated subbands.

6.7 Dyadic Sorting Routine

To ensure a uniform permutation between all subbands for $p = 0, 1, \dots, \tilde{M} - 1$, the detection and correction criterion described in Section 6.6 must be applied between all pairs of separated subbands. The easiest approach to do this is via a sequential sorting routine where we correct the permutation for the first two subbands and then every adjacent subband after that relative to it's previous subband. However as shown in (K. Rahbar, and J. Reilly, 2001), the disadvantage of this is that in a worse case scenario, half of all the subbands will have one permutation and the other half

will have another. To prevent this we adopt the same dyadic tree-structured sorting routine as in (K. Rahbar, and J. Reilly, 2001) however we now do it with respect to separated subbands as opposed to frequency bins and also assume that the total number of subbands \tilde{M} is an integer power of 2. An example of such a tree structure is shown in Figure 6.4. After detection and correction of permutations in adjacent subbands, corrected subband signals belonging to each individual sorted group are added together to obtain the next signal corresponding to higher level up in the tree. Again the subband permutation detection criterion is established using the envelopes of the new signals at the higher level. This process is repeated until all subbands have the same permutation at the lowest level of the tree. A brief description of the sorting routine is given below.

1. Assume a hierarchial level index l , where $l = 0, 1, \dots, \log_2(\tilde{M}) - 1$. The zeroth level is the bottom level in Figure 6.4 and is also the initial value for l . Also define $\Sigma_p^0(k) = \mathbf{R}_{\tilde{s}\tilde{s},k}^{\tau,p}$, where $p = 0, 1, \dots, \tilde{M} - 1$ and $\tau = 0$.

FOR $l = 0, 1, \dots, \log_2(\tilde{M}) - 1$, START

2. Divide the available separated subbands $\hat{\mathbf{s}}_{\{1,2,\dots,N\}}^{sub}(p,k)$ at hierarchial level l into groups of two bins, with group index u , where $u = 0, 1, \dots, \tilde{M}/(2^{l+1}) - 1$.
3. Let $\mathbf{\Pi}^l(\hat{\mathbf{s}}_j)$ for each $j \in [2u, 2u + 1]$ be the permutation matrix for the hierarchial level l and group u estimated from the proposed permutation criterion. Then for all groups $u = 0, 1, \dots, \tilde{M}/(2^{l+1})$ and $j \in [2u, 2u + 1]$, we update the order of diagonal values of $\Sigma_j^0(k)$ using

$$\Sigma_j^l(k) = \mathbf{\Pi}^l(\hat{\mathbf{s}}_j) \Sigma_j^l(k) \mathbf{\Pi}^l(\hat{\mathbf{s}}_j)^T. \quad (6.22)$$

We only update the diagonal values of $\Sigma_j^0(k)$ for one value of j in the set $[2u, 2u + 1]$ so we do not do redundant permutation sorting.

4. Update the order of the columns of $\hat{\mathbf{s}}_{\{1,2,\dots,N\}}^{sub}(j, k)$ using

$$\hat{\mathbf{s}}_{\{1,2,\dots,N\}}^{sub}(j, k) \mathbf{\Pi}^l(\hat{\mathbf{s}}_j) \quad (6.23)$$

5. For each group u calculate

$$\Sigma_u^{l+1}(k) = \Sigma_{2u}^l(k) + \Sigma_{2u+1}^l(k), \quad l \neq \log_2(\tilde{M}) - 1 \quad (6.24)$$

END

6.8 Simulation Results

6.8.1 Benchmark ICA'99 Dataset

There are numerous methods in the field of blind signal separation used to evaluate the performance of various separation algorithms including plots of separated signals, plots of cascaded mixing/demixing impulse responses and signal to noise

ratios. It is important to use standard data test sets that are available to provide a unified methodology to making a good comparative analysis between algorithms in an objective manner. Controllable synthetic test cases are used to examine algorithm performance in trivial to moderately complex test cases allowing accurate evaluation of separation for different algorithms where information of sources and mixing/demixing systems are available. In comparison however it is also equally important to test the algorithms in a real environment to demonstrate the effectiveness of the algorithm in an application sense. Real world recordings for acoustic signal separation should be considered to reflect the complexity of real mixing systems and the success of separation for different BSS algorithms. Sources, mixing systems and performance measures for synthetic and real cases that are standard tools for evaluating blind signal separation are referred to in (D. Schobben, K. Torkkola, and P. Smaragdis, 1999).

6.8.2 Synthetic Testing

To test the proposed cosine subband convolutive BSS algorithm with the proposed permutation detection and dyadic sorting network we use the synthetic mixing impulse response filters for a TITO network generated in a virtual room with dimensions $10m \times 10m \times 10m$. The impulse responses of the mixing system with reference to the positions of the sources and microphones in the virtual room, shown in Figure 6.5, are generated using the *simroommix.m* function available from <http://www2.ele.tue.nl/ica99>. Assuming a sampling rate of 8 kHz, the synthetic impulse responses obtained are shown in Figure 6.6. They have a reverberation time of 130 ms corresponding to a filter length of $P = 1039$. The primary or dominant echo information of the virtual room is found in the first 64 ms of the impulse responses, corresponding to $P = 512$ so we ignore the trailing information. Using the Wiener filtering approach to obtain the corresponding demixing filters explained in Chapter 3, we obtain demixing FIR

Virtual Room Setup

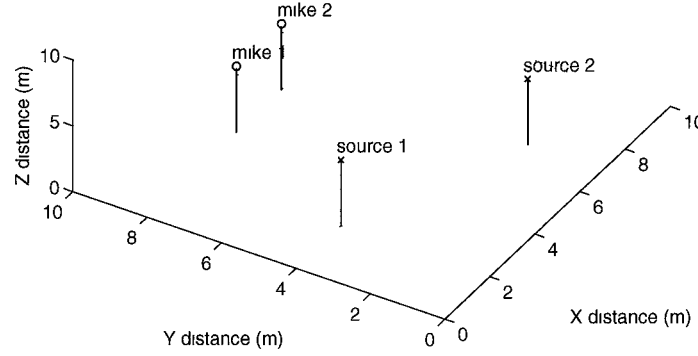


Figure 6.5 Virtual room synthetic mixing environment at 8kHz sampling frequency generated with the `simroommix.m` function

filters with impulse responses shown in Figure 6.7. These have a reverberation time of 200 ms corresponding to a filter length of $Q = 1600$. The mixing and demixing TITO FIR systems in cascade produces a global system which is a scaled and delayed version of the identity matrix \mathbf{I}_N . The delay from the Wiener solution corresponds to a delay of 32 ms.

With the TITO fullband demixing system available, we then employ the method proposed in (J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, 2002) which allows us to obtain the correct subband components for each of the respective demixing channels. Subband decomposition of a large filter into its subband components via a uniform \tilde{M} channel CM FIR FB with a subsampling factor of R is given via the following least-squares approximation taken from (J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, 2002) and adapted for our problem,

$$\mathbf{w}_{(l,j)_p LS} = \mathbf{H}_p^\dagger \bar{\mathbf{u}}(n)_{p \downarrow R}, \quad (6.25)$$

where $p = 0, 1, \dots, \tilde{M} - 1$ is the subband index, \mathbf{H}_p^\dagger is the Moore-Penrose pseudo-

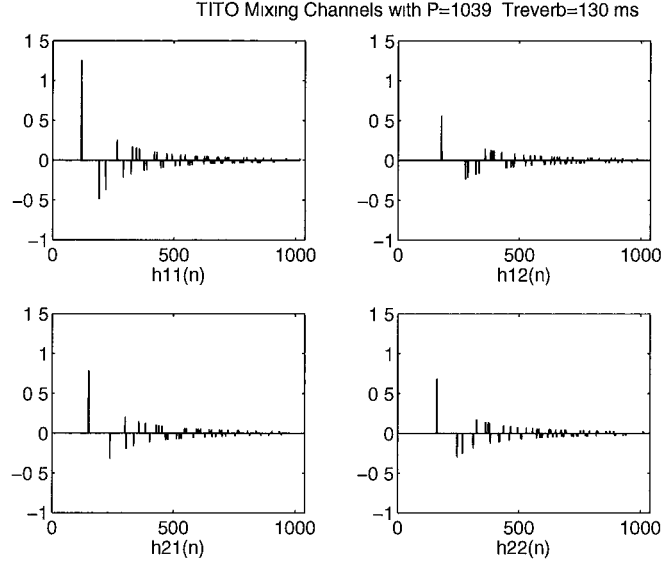


Figure 6.6 Virtual room synthetic mixing impulse responses at 8kHz sampling frequency and reverberation time of 130 ms.

inverse of the Toeplitz convolutive matrix \mathbf{H}_p defined in (J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, 2002), and $\bar{\mathbf{u}}(n)_{p \downarrow R} \triangleq (h_p(n) * \mathbf{w}_{i,j}(n))_{\downarrow R}$. The length of each of the subband components for each respective fullband demixing impulse responses can be defined as,

$$Q_p = \lceil \frac{Q + L - 1}{R} \rceil - \lceil \frac{L}{R} \rceil + 1 \quad (6.26)$$

By having an idea of what the subband components for each demixing FIR filter in the TITO system should be, a comparison can be made to blind solutions for the separating TITO system in each subband after uninitialized global optimization to see whether the scaling and permutation indeterminacies are evident between separated subbands. For the synthetic case it was assumed that there were $\tilde{M} = 256$ subband channels, and the oversampling ratio was 2, giving $R = 128$. The length of the analysis FIR filters was $L = 4\tilde{M} = 1024$ and the length of the fullband demixing FIR filters was $Q = 1600$, resulting in a length of subband components for each demixing TITO system of $Q_p = 14$. So now instead of having a single multivariate non-convex

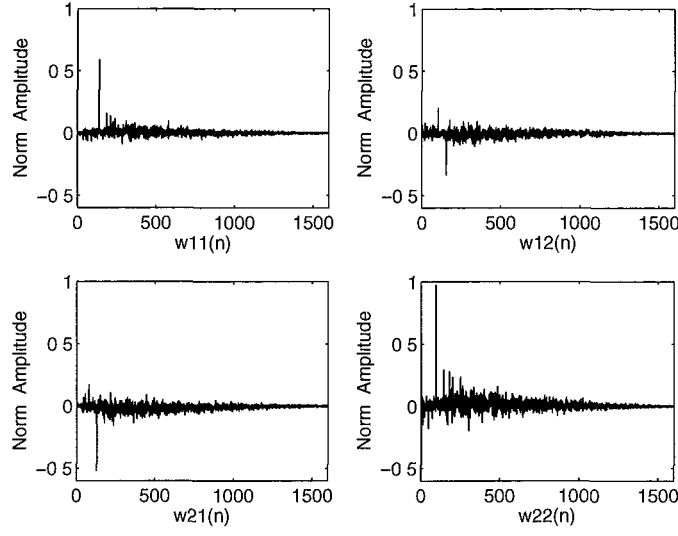


Figure 6.7 Wiener solution to TITO FIR synthetic demixing system with reverberation time of 200ms i.e. $Q=1600$, delay=32ms

non-linear optimization problem with 6400 dimensions for a TITO demixing system, there is $p = \tilde{M} = 256$ optimization problems each with 56 dimensions to solve for. The two input signals $s(t)$ were 4 second speech segments taken from two different utterances from the TIMIT corpus of speech. The signals were matched to ensure the average volume level was the same and then were downsampled to a sampling rate of 8 kHz. The input signals were mixed with the synthetic mixing filters $\mathbf{H}(\tau)$, taken from the virtual room setup in Figure 6.5, to generate the fullband observed signals to be used as input into the CM FIR FB shown in Figure 6.1. Assuming the input signal is stationary over a specific time frame then for speech this time frame is a period of 20 ms. At a sampling rate of 8 kHz this corresponds to frame of length 160 samples. The observed signal is then segmented into frames of length 20 ms each and a subband decomposition of every frame of the observed signal is done to obtain the cross-correlation matrices $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^p(\tau)$, for every stationary frame k , over all time lags τ , to carry out the joint diagonalization task using global optimization for each respective subband p individually. Figure 6.9 shows the separated subbands for

the first 2 adjacent subbands. It was noted that in this particular simulation there was no scaling discrepancy between separated subbands, however there was permutation indeterminacies at various subbands. After the separation stage the first non-zero k th stationary frame from each respective separated subband was checked for any permutation using the envelope correlation criteria described in Section 6.6. Detection and correction of permutations between all adjacent subbands was carried out using the dyadic sorting routine to ensure consistent results across all pairs of adjacent subbands. The separated subbands were passed through the synthesis stage of the CM FIR FB to obtain the fullband separated signals. A qualitative analysis was done with listening tests of the separated signals indicating fairly good separation. In addition to this a quantitative analysis was done by comparing the signal-to-interference (SIR) ratio of our proposed method with a typical frequency domain approach described in Section 6.2 from (L. Parra and C. Spence, 2000) with geometric beamforming used for initialization from (L. Parra and C. Alvino, 2002). The angle of incidence or direction of arrival (DOA) of the beam was calculated using simple trigonometry from the location of the sources and sensors of the virtual room in Figure 3.4. The SIR is defined as $SIR = SIR_O - SIR_I$, where,

$$SIR_O = 10 \log \frac{\sum_{\omega} \sum_i |A_i(\omega) S_i(\omega)|^2}{\sum_{\omega} \sum_{i \neq j} \sum_j |A_{ij}(\omega) S_j(\omega)|^2}, \quad (6.27)$$

$$SIR_I = 10 \log \frac{\sum_{\omega} \sum_i |H_i(\omega) S_i(\omega)|^2}{\sum_{\omega} \sum_{i \neq j} \sum_j |H_{ij}(\omega) S_j(\omega)|^2}. \quad (6.28)$$

$\mathbf{A}(\omega) = \mathbf{W}(\omega)\mathbf{H}(\omega)$ and $i \neq j$. SIR means the ratio of a target-originated signal to a jammer-originated signal (L. Parra and C. Spence, 2000). The SIR metric above can be used for the frequency domain approach however to use it for the subband domain approach there must be a method to convert from the subband demixing components back to the fullband demixing system. As we do not currently have a method of doing this we calculate the equivalent technique in the subband domain for finding the SIR. In the subband approach, the SIR is defined as $SIR = SIR_O - SIR_I$,

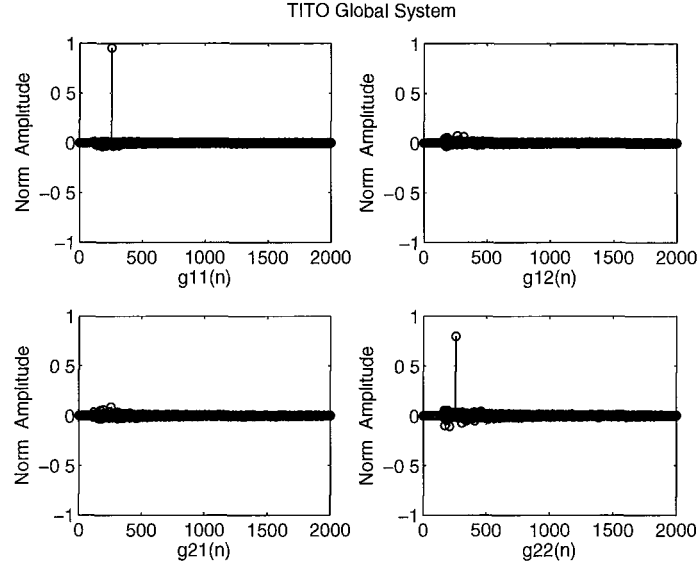


Figure 6.8 Ideal global TITO FIR system for virtual room.

where,

$$SIR_O = 10 \log \frac{\sum_p \sum_i |A_{ip} * S_{ip}|^2}{\sum_p \sum_{i \neq j} \sum_j |A_{ijp} * S_{jp}|^2}, \quad (6.29)$$

$$SIR_I = 10 \log \frac{\sum_p \sum_i |H_{ip} * S_{ip}|^2}{\sum_p \sum_{i \neq j} \sum_j |H_{ijp} * S_{jp}|^2}. \quad (6.30)$$

$\mathbf{A}_p = \mathbf{W}_p * \mathbf{H}_p$ and $i \neq j$. Note that the subband components of the fullband mixing system can be found using the method described in (J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, 2002). The typical BSS frequency domain approach sets the required variables of the algorithm to be $Q = 1600, T = 4096, K = 204$ and an initial value for the unknown system in each frequency bin T that is derived using prior knowledge of the geometrical location of the sources respective to the sensors as described in (L. Parra and C. Alvino, 2002). Table 6.1 shows that the SIR was slightly better for our proposed approach without the need for initialization when performing the optimization of the objective function.

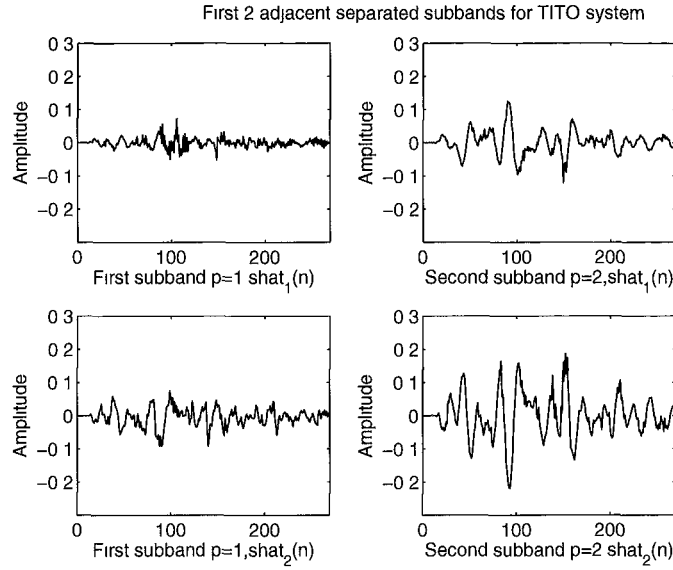


Figure 6.9 First two separated adjacent subbands for $p = 1, 2$ when $\tilde{M} = 256$, $R = 128$.

Table 6.1 Separation performance comparison using SIR

Alg.	SIR (dB) Synth.	SIR (dB) Real
Parra/Spence	18.2	15.4
SBSS	21.3	17.5

6.8.3 Real Testing

To test a real case scenario, the same approach for the synthetic testing model was used but instead of using artificially produced mixing impulse responses, recordings were taken in a real reverberant room. In Chapter 3, the process of generating the impulse responses taken from a reverberant environment was explained. The benefit of this is that we can work out a good approximation of the mixing system, and then find the corresponding MIMO fullband inverting demixing system and its corresponding subband components for measures of reference to our blind solution in the subband domain. The measured room impulse response was found using the MLS method as explained previously. The TITO mixing system had a reverberation time of 200ms corresponding to FIR filters of length $P = 1600$ shown in Figure 3.5. The demixing

FIR filters found by solving the Wiener-Hopf equations had a reverberation time of 250ms corresponding to FIR filters of length $Q = 2000$. Using the method proposed in (J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, 2002), we found the subband components using the least-squares approach for the demixing FIR filters of the fullband TITO demixing system. Using a CM FIR FB with $\tilde{M} = 512$ and $R = 256$, Equation (6.26) gives the length of the unknown demixing systems \mathcal{W}_p to be $Q_p = 9$ for all subbands. The same two input speech signals of 4 seconds in length used in the synthetic testing and taken from the TIMIT corpus of speech were used in this simulation. At a sampling frequency of 8 kHz, the number of stationary frames to jointly diagonalize in each subband was $K = 210$. After detecting and correcting permutations between adjacent subbands using the dyadic sorting approach, the separated subbands were passed through the synthesis stage of the filter bank to obtain the fullband separated signals up to a global permutation and scaling factor. The quality of separation from this method was also compared to the frequency domain method described in Section 6.2 which uses an initialization based on *a priori* knowledge using geometric beamforming taken from (L. Parra and C. Alvino, 2002). The parameters for the frequency domain method in this case were $Q = 2000$, $T = 4096$, and $K = 210$ with an angle of incidence calculated from Figure 3.4. The SIR for both methods is given in Table 6.1 showing a slight improvement without the extra knowledge of the location of sources to sensors required for the initialized frequency domain approach. With a slightly larger demixing system to solve for due to a longer reverberation time, the SIR is slightly lower for both approaches than the synthetic case as shown in Table 6.1. For better performance it is expected that the effects of aliasing due to the subband processing introduced can be decreased by increasing the oversampling factor although this will also have implications on the size of the unknown demixing systems within each subband.

6.9 Conclusions

The main contributions of this chapter collectively demonstrate a general framework to approach the problem of BSS using a subband approach when convolutive mixing of nonstationary audio/speech sources in a reverberant environment exists. The motivation of using a subband analysis in the context of BSS as opposed to the typical frequency domain approach has been briefly discussed due to the upper bounds and limitations on separation performance for the frequency domain approach as discussed in (S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, 2001; M. Ikram, and D. Morgan, 2000; R. Mukai, S. Araki, H. Sawada, and S. Makino, 2004) and verified by our SIR results for our proposed subband method. A methodology for conducting experiments to obtain room impulse responses, their corresponding inverting systems and the subband components for MIMO systems has been identified and serves to provide available information in the *non-blind* case to compare *blind* results to. The introduction of global optimization as a method to solve blind systems without the need for additional information about source/sensor locations is investigated using a branch-and-bound technique and can be used for smaller type BSS convolutive problems without the frequency permutation problem. For larger reverberant blind systems, subband decomposition allows large problems with infeasible convergence times to be reduced to many smaller type BSS problems. Avenues for future research include investigation of other PR filter bank polyphase models and global optimization methods that can improve separation performance further whilst simultaneously reducing computational complexity and convergence times.

Chapter 7

Conclusions and Suggestions for Further Research

7.1 Conclusions

The focus of this thesis has aimed at developing a methodology and framework to model the convolutively mixed BSS problem in the subband domain, with an emphasis on applications involving reverberant acoustic environments and nonstationary signals such as speech. The validity of proposing such a subband model to conduct BSS for convolutively mixed signals as opposed to the typical frequency domain approach is justified through the limitations and upper bounds placed on separation performance in longer reverberant environments for the frequency domain approach as evidenced in (L. Parra and C. Alvino, 2002; H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004a) and (H. Sawada, R. Mukai, S. Araki, and S. Makino, 2004b).

Firstly as proposed in (Chapter 3), to assess the performance of the model developed, *a priori* information available in a non-blind sense aids in a comparative analysis of the proposed model with reference to other typical models/approaches to solving the same problem. Developing a systematic approach to obtaining all relevant information of the problem space is paramount and provides a more analytical framework for

comparison of the proposed subband model to a typical frequency domain approach. The method proposed serves as a tool for conducting a more effective comparative analysis of separation quality between two or more BSS approaches to the problem of separation of speech in a reverberant acoustic environment. A framework for obtaining the impulse responses of a MIMO system used to model a room response is proposed using the MLS technique. Then the process of obtaining the inverse response of the MIMO system is proposed utilizing the Wiener-Hopf theory to obtain the appropriate FIR filter responses of the corresponding MIMO demixing system is proposed. This provides all relevant information of the convolutive BSS model, realized for a practical problem.

For all convolutive BSS problems, there is always the inherent issue of the local frequency permutation problem. Many typical approaches transform the problem into the frequency domain, via the DFT, and attempt to solve many instantaneous BSS problems in each frequency bin. In (Chapter 4), we proposed two new algorithms in time and in time-frequency domains to avoid the local permutation problem. There are implications of solving the local permutation problem in the frequency domain. Usually the benefits of transforming a convolutive time-domain problem to a multiplicative frequency domain problem is evident due to savings in computations and convergence time, which is good for practical real time applications. However, the benefits of proposing a pure time-domain or time-frequency domain system become evident for small to medium size convolutive mixing systems as the savings of not having to solve the local permutation problem, outweigh the costs of performing the DFT transform twice, and having to use some projection operation to avoid the local permutation problem, as is the case with typical approaches in literature. This chapter proposes two new BSS algorithms that avoid the local permutation problem that remains evident in other approaches.

Most current BSS models that are used to solve convolutive BSS problems with non-stationary sources such as speech make use of some additional information or assumption of the problem and use local optimization routines. For example, with speech signal processing and array/geometric beamforming for BSS in a room, i.e. the 'cocktail party' problem, prior knowledge of the spatial layout of sources with respect to the sensors is used to provide good initial starting points for the unknown demixing system so as to avoid ill-convergence to local multim minima in a non-convex multimodal optimization problem. In (Chapter 5), we propose the integration of a specific branch-and-bound global optimization method with the convolutive BSS time domain model proposed in (Chapter 4), to solve the BSS problem without reference to any additional information or assumptions. This makes the problem truly blind which in essence is the idea behind BSS. Where there exist problems that one cannot assume additional criteria on the problem, the justification of such a proposed method becomes evident. The DiRect algorithm is customized to fit the BSS problem criteria for convolutive mixing and global optimization is performed and compared to existing techniques.

For (Chapters 3,4, and 5), the main benefits are primarily related to smaller to medium sized convolutive mixing systems. In practical problems however where we have reverberant environments doing the mixing, impulse responses for MIMO FIR filter systems become very long. In such cases it becomes more viable to conduct signal separation in the frequency or subband domain. With the limitations on frequency domain signal separation performance being cited previously in this thesis and discussed in Chapter 2, the motivation to develop a subband based framework to conduct BSS in for convolutive models becomes apparent. In (Chapter 6) we integrate the different proposed models and methods of this thesis into a subband framework. Subband decomposition is performed on observed signals using oversampled CM

and DFT FIR filter bank models based on an ELT prototype. Both of these models are compared when performing initialized BSS with the Newton local optimization method, and then uninitialized BSS with the DiRect global algorithm within each subband is performed and compared to the typical frequency domain approach. For the global approach it is seen that a CM FIR filter bank subband decomposition, followed by the separation phase performs slightly better than a typical frequency domain separation approach as far as quality of separation is concerned. Separation in the subband domain, like the frequency domain, requires the solving of the local permutation problem, and a new detection of permuted separated adjacent subbands is proposed along with a dyadic sorting routine to ensure correct alignment of all separated subbands over the entire spectrum.

7.2 Suggestions for Future Research

A potentially fruitful area of future research work is in applying the uninitialized global optimization methods for convolutive time domain BSS with nonstationary source signals to wireless communications applications, where there exist more smaller to medium scale mixing and demixing systems which can benefit from the proposed algorithms presented in (Chapters 4 and 5). For those communications applications where the observed signals are represented as complex signals, the DiRect global optimization algorithm must be changed slightly to handle optimization of multivariate complex signals. Also, if DFT composition is used on real observed signals resulting in complex frequency bin signals, then the DiRect global optimization algorithm again would need to be changed slightly to accommodate the complex case.

The proposed subband detection and correction schemes currently only consider the TITO mixing and demixing system case. The proposed method needs to be extended

to accommodate the more general MIMO mixing and demixing system cases. Also, the current structure of FIR filter bank used is a direct one. As explained in Section 2.9, polyphase representation and design of filter banks would prove more computationally beneficial resulting in faster implementations. In (Malvar, 1992) presented a fast algorithm for any overlapping factor using orthogonal butterfly angles and a type-IV DCT. With reference to this, the subband FIR filter bank could be improved. In addition to this, a further investigation into the filter bank parameters including the overlapping factor, the design of the prototype window, the method of modulation, the subsampling factor, and the number of subbands is suggested.

Different global optimization methods could also be considered in relation to optimizing the BSS problem when there is no additional information present in the problem space, within the subband domain.

Bibliography

- A. Bell, and T. Sejnowski (1995). Acoustics, speech, and signal processing. volume 5, pages 3415–3418. ICASSP.
- A. Hyvärinen, J. Karhunen, and E. Oja (2001). *Independent Component Analysis*. John Wiley Sons, Inc., New York, NY.
- A. Liavas, P. Regalia, and J. Delmas (1999). Blind channel approximation: Effective channel order determination. *IEEE Transactions on Signal Processing*, 47(12):3336–3344.
- A. Mertins (1999). *Signal Analysis*. John Wiley and Sons Ltd.
- A. Westner, and V. Bove (1999). Blind separation of real world audio signals using overdetermined mixtures. Aussois, France. International Workshop on Independent Component Analysis and Signal Separation.
- A. Yeredor (2002). Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Trans., Signal Processing*, 50:1545–1553.
- B. Krongold and D. Jones (2000). Blind source separation of nonstationary convolutive mixed signals. pages 53–57, Pocono Manor, PA. Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing.

- C. Chang, Z. Ding, S. Yau, and F. Chan (2000). A Matrix-Pencil Approach to Blind Separation of Colored Nonstationary Signals. *IEEE Trans., on Signal Processing*, 48(3):900–907.
- Cichocki A. and Amari S.-I. (2002). *Adaptive Blind Signal and Image Processing*. Wiley, Chichester.
- D. Pham, and J. Cardoso (2001). Blind separation of instantaneous mixtures of non-stationary sources. *IEEE Transactions on Signal Processing*, 49:1837–1848.
- D. R. Jones, C. D. Perttunen, and B. E. Stuckman (1992). Global Optimization: Beyond the Lipschitzian Model. pages 566–570. IEEE International Conference on Systems, Man and Cybernetics.
- D. R. Jones, DIRECT (2001). *Encyclopedia of Optimization*. Kluwer Academic Publishers.
- D. Schobben, K. Torkkola, and P. Smaragdis (1999). Evaluation of blind signal separation methods. pages 261–266, Aussois, France. International Workshop on Independent Component Analysis and Signal Separation.
- D. Yellin, and E. Weinstein (1994). Criteria for multichannel signal separation. *IEEE Transactions on Signal Processing*, 42:2158–2168.
- F. Ehlers, H. Schuster (1997). Blind separation of convolutive mixtures and an application in automatic speech recognition in a noisy environment. *IEEE Transactions on Signal Processing*, 45:2608–2612.
- F. Preparata, and M. Shamos (1985). *Computational geometry: an introduction*, pages 108–109. Springer-Verlag New York, Inc.
- Feng, M. and Kammeyer, K. (1998). Blind Source separation for Communication Signals Using Antenna Arrays. Florence, Italy. accepted by ICUPC.

- Gill, P. (2002). User's guide for snopt v6, A Fortran package for large-scale nonlinear programming. URL: <http://tomlab.biz/docs/snoptA.pdf>.
- H. Malvar (1990). Modulated QMF filter banks with perfect reconstruction. *Electronics Letters*, 26:906–907.
- H. Malvar (1991). Extended Lapped Transforms: fast algorithms and applications. pages 1797–1800, Toronto, Canada. IEEE Int. Conference on Acoustics, Speech, and Signal Processing.
- H. Malvar (1992). Extended Lapped Transforms: properties, applications, and fast algorithms. *IEEE Trans., on Signal Processing*, 40(11):2703–2714.
- H. Nussbaumer (1981). Pseudo-QMF filter bank. volume 24, pages 3081–3087. IBM Tech. Disclosure Bull.
- H. Sawada, R. Mukai, S. Araki, and S. Makino (2004a). A robust and precise method for solving the permutation problem of frequency-domain blind source separation. *IEEE Trans., on Speech and Audio Processing*, 12(5):530–538.
- H. Sawada, R. Mukai, S. Araki, and S. Makino (2004b). Convolutional blind source separation for more than two sources in the frequency domain. pages 885–888, Montreal, Canada. IEEE Int., Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004).
- Holmström, K. (2002). User's Guide for TOMLAB v4.0. URL: <http://tomlab.biz/docs/tomlabv4.pdf>.
- Ikeda, S. and Murata, N. (1999). A method of ICA in time-frequency domain. In *Proc. ICA*, pages 365–361, Aussois.
- J. Anemüller (2001). *Across-frequency processing in convolutional blind source separation*. PhD thesis, University of Oldenburg.

-
- J. Anemuller, and B. Kollmeier (2000). Amplitude modulation decorrelation for convolutive blind source separation. pages 215–220, Helsinki, Finland. 2nd IEEE Int. Workshop on Independent Component Analysis.
- J. Borish, and J. Angell (1983). An efficient algorithm for measuring the impulse response using pseudorandom white noise. *Journal of Audio Eng. Society*, 31(7/8):478–488.
- J. Cardoso (1998). Blind Signal Separation: Statistical Principles. volume 86, pages 2009–2025. Proceedings of the IEEE.
- J. Choi (2004). Equalization and Semi-Blind Channel Estimation for Space-Time Block Coded Signals Over a Frequency-Selective Fading Channel. *IEEE Transactions on Signal Processing*, 52(3):774–785.
- J. Herault, and C. Jutten (1986). *Space or time adaptive signal processing by neural network models*, volume 151, chapter in Neural Networks for Computing, pages 206–211. AIP press.
- J. Klierwer, and A. Mertins (1998). Oversampled Cosine-Modulated Filter Banks with Arbitrary System Delay. *IEEE Trans., on Signal Processing*, 46(4):941–955.
- J. Manton (2002). Optimization exploiting unitary constraints. *IEEE Trans. on Signal Processing*, 50(3):635–650.
- J. McClellan, R. Schafer, and M. Yodar (1993). *DSP First*. Prentice Hall, Inc, Upper Saddle River, NJ.
- J. Mendal (1990). *Maximum-Likelihood deconvolution*. Springer-Verlag New York Inc., New York, NY.

- J. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand (2002). The complex subband decomposition and its application to the decimation of large adaptive filtering problems. *IEEE Trans., on Signal Processing*, 50(11):2730–2743.
- J. W. Brewer (1978). Kronecker products and matrix calculus in system theory. *IEEE Trans. Circuits Syst.*, 25:772–781.
- K. Holmström, and M. Edvall (2004). glcCluster - A combined DIRECT, clustering and local search algorithm for global nonconvex optimization. URL: <http://www.tomlab.biz/download/publications.php>.
- K. Pope and R. Bogner (1996a). Blind signal separation I: Linear, instantaneous combinations. *Digital Signal Processing*, 6:5–16.
- K. Pope and R. Bogner (1996b). Blind signal separation II: Linear, convolutive combinations. *Digital Signal Processing*, 6:17–28.
- K. Rahbar, and J. Reilly (2001). Blind source separation algorithm for MIMO convolutive mixtures. pages 242–247, San Diego, CA.
- K. Rahbar, J. Reilly, and J. Manton (2002). A frequency domain approach to blind identification of MIMO FIR systems driven by quasi-stationary signals. volume 2, pages 1717–1720, Orlando, Florida, USA. IEEE International Conference on Acoustics, Speech, Signal Processing.
- K. Rahbar, J. Reilly and J. Manton (2004). Blind Identification of MIMO FIR Systems Driven by Quasistationary Sources Using Second-Order Statistics: A Frequency Domain Approach. *IEEE Trans., on Signal Processing*, 52(2):406–417.
- K. Torkkola (1999). Blind separation of audio signals - Are we there yet? pages 239–244. Proceedings of ICA and Signal Separation.

- Koilpillai, R. and Vaidyanathan, P. (1992). Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction. *IEEE Trans. on Signal Processing*, 40:770–783.
- L. Parra and C. Alvino (2002). Geometric Source Separation: Merging Convolutional Source Separation With Geometric Beamforming. *IEEE Trans., on Speech and Audio Processing*, 10(6):352–362.
- L. Parra and C. Spence (2000). Convolutional Blind Separation of Non-Stationary Sources. *IEEE Trans., Speech and Audio Proc.*, 8:320–327.
- L. Parra, C. Spence, and B. Vries (1997). Convolutional source separation and signal modeling with ML. Reggio Calabria, Italy. International Symposium on Intelligent Systems (ISIS'97).
- Larsen, J., Hansen, L., Kolenda, T., and Nielsen, F. (2003). Independent Component Analysis in Multimedia Modelling. pages 687–696, Nara, Japan. ICA.
- M. Björkman and K. Holmström (1999). Global Optimization with the DIRECT Algorithm in Matlab. *Advanced Modeling and Optimization*, 1(2):17–37.
- M. Feder, and E. Weinstein (1988). Parameter-estimation of super-imposed signals using the EM algorithm. *IEEE Trans., on Acoustics, Speech, and Signal Processing*, 36:477–489.
- M. Hofbauer (2004a). On the FIR inversion of an acoustical convolutional mixing system: properties and limitations. volume 3195, pages 643–743, Granada, Spain. Springer-Verlag GmbH.
- M. Hofbauer (2004b). On the FIR inversion of an acoustical convolutional mixing system: properties and limitations. Granada, Spain. ICA.

- M. Ikram, and D. Morgan (2000). Exploring permutation inconsistency in blind separation of signals in a reverberant environment. pages 1041–1044, Istanbul, Turkey. Proceedings of the IEEE Int. Conf. on acoust., Speech and Signal Processing.
- M. Ikram, and D. Morgan (2002). A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. pages 881–884. Proc ICASSP.
- M. Joho (2004). Blind Signal Separation of convolutive mixtures: A time domain joint-diagonalization approach. pages 578–585, Granada, Spain. ICA.
- M. Joho and K. Rahbar (2002). Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation. *Sensor Array and Multichannel Signal Processing Workshop Proceedings*, pages 403–407.
- Malvar, H. S. (1992). *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA.
- N. Murata, S. Ikeda, and A. Ziehe (2001). An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, 41, Issue 1-4:1–24.
- N. Sellami, M. Siala, and I. Fijalkow (2004). Low-complexity equalizers for mimo frequency selective channels. pages 175–178. First International Symposium on, Control, Communications and Signal Processing.
- P. Comon, C. Jutten, and J. Herault (1991). Blind separation of sources, Part II: problems statement. *Signal Processing*, 24:11–20.
- P. E. Gill, W. Murray, and M. A. Saunders (1997). SNOPT: An SQP algorithm for large-scale constrained optimization. Technical report, Numerical Analysis

- Report 97-2, Dept. of Mathematics, University of California, San Diego, La Jolla, CA.
- P. Vaidyanathan (1993). *Multirate systems and filter banks*. Prentice Hall, Inc., Englewood Cliffs, NJ.
- Parra, L. and Spence, C. (2000). On-line blind source separation of non-stationary signals. *Journal of VLSI Signal Processing*, 26(1/2):39–46.
- R. Aichner, and H. Buchner (2003). On-line time-domain blind source separation of nonstationary convolved signals. Nara, Japan. Int. Symposium on Independent Component Analysis and Blind Signal Separation.
- R. Fletcher, and S. Leyffer (1997). Nonlinear programming without a penalty function. Technical report, University of Dundee, NA/171.
- R. Horst, and P. M. Pardalos (1995). *Handbook of Global Optimization (Nonconvex Optimization and Its Applications)*, volume 1. Kluwer.
- R. Lambert (1996). *Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures*. PhD thesis, University of Southern California.
- R. Lambert, and A. Bell (1997). Blind separation of multiple speakers in a multipath environment. Munich, Germany. ICASSP.
- R. Mukai, H. Sawada, S. Araki, and S. Makino (2004). Frequency domain blind source separation for many speech signals. pages 461–469, Granada, Spain. ICA.
- R. Mukai, S. Araki, H. Sawada, and S. Makino (2004). Evaluation of separation and dereverberation performance in frequency domain blind source separation. *Acoustical Science and Technology*, 25(2):119–126.

- R. Rajagopal (2000a). Exact FIR Inverses of FIR Filters. Master's thesis, Graduate School Ohio State University, Columbus, Ohio.
- R. Rajagopal (2000b). Exact FIR inverses of FIR filters. Master's thesis, Graduate School of The Ohio State University.
- Rahbar, K. and Reilly, J. (2003). A new frequency domain method for blind source separation of convolutive audio mixtures. *Submitted to IEEE Trans. on Speech and Audio Processing*.
- S. Amari (1998). Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276.
- S. Araki, S. Makino, R. Aichner, T. Nishikawa and H. Sarawatari (2003). Subband based blind source separation with appropriate processing for each frequency band. pages 499–504, Nara, Japan. 4th Int. Sym. on ICA and BSS.
- S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari (2001). Fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. pages 2737–2740, Salt Lake City, Utah. Proceedings of the IEEE Int. Conf. on Acoust., Speech and Signal Processing.
- S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura (2000). Evaluation of blind signal separation method using directivity pattern under reverberation conditions. pages 3140–3143. Proc. ICASSP.
- S. Makeig, A. Bell, T. Jung, and T. Sejnowski (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Proc. Systems*, (8):145–151.
- Shure Incorporated (2004). Model SM57 dynamic microphone. URL: http://www.shure.com/pdf/specsheets/spec_wiredmics/sm57.pdf.

- T. Petermann, D. Boss, and K. Kammeyer (1999). Blind GSM Channel Estimation Under Channel Coding Conditions. pages 180–185, Phoenix, USA.
- W. Wang, J. Chambers, and S. Sanei (2004). A novel hybrid approach to the permutation problem of frequency domain blind source separation. pages 532–539, Granda, Spain. ICA.

Appendix A

Proof of Closed Form Analytical Expressions for Gradient and Hessian

A.1 Proof of Closed Form Analytical Expressions for Gradient and Hessian

Useful equations and relations for matrix calculus, Frobenius norm, trace, $\text{vec}(\cdot)$, and the Kronecker product functions, originally cited from (J. W. Brewer, 1978) and also given in appendices B and C in (M. Joho and K. Rahbar, 2002) are helpful in the following derivations. We firstly provide the derivation of the objective function \mathcal{J}_3 , and then the constraint \mathcal{J}_4 . The objective function is given as

$$\mathcal{J}_3(\mathcal{W}) = \sum_{\tau=-\tau_{m_n}}^{\tau_{max}} \sum_{k=1}^K \beta_k \tau \|off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}}^{\tau} \mathcal{W}^H)\|_F^2 \quad (\text{A } 1)$$

If we ignore the summations we get,

$$\begin{aligned} \mathcal{J}_3(\mathcal{W}) &\triangleq \|off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)\|_F^2 \\ &= \|\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H\|_F^2 - \|diag(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)\|_F^2 \\ &= tr(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H \mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \\ &\quad - tr(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H diag(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) \end{aligned} \quad (\text{A } 2)$$

Following on from (J. Manton, 2002) and (M. Joho and K. Rahbar, 2002), the second-order Taylor series approximation of some cost function \mathcal{J} with respect to $\mathcal{W} \in \mathbb{C}^{N \times M_Q}$ in the non-square matrix form is,

$$\begin{aligned} \mathcal{J}(\mathcal{W} + \delta \mathbf{Z}) &= \mathcal{J}(\mathcal{W}) + \delta \mathbb{R}\{tr(\mathbf{Z}^H \mathbf{G}_{\mathcal{W}})\} \\ &\quad + \frac{\delta^2}{2} \text{vec}(\mathbf{Z})^H \mathbf{H}_{\mathcal{W}} \text{vec}(\mathbf{Z}) \\ &\quad + \frac{\delta^2}{2} \mathbb{R}\{\text{vec}(\mathbf{Z})^T \mathbf{C}_{\mathcal{W}} \text{vec}(\mathbf{Z})\} + O(\delta^3), \end{aligned} \quad (\text{A } 3)$$

where $\mathbf{G}_{\mathcal{W}} \in \mathbb{C}^{N \times M_Q}$ is the derivative of \mathcal{J} evaluated at \mathcal{W} , and $\{\mathbf{H}_{\mathcal{W}} + \mathbf{C}_{\mathcal{W}}\} \in \mathbb{C}^{NM_Q \times NM_Q}$ is the Hessian of \mathcal{J} evaluated at \mathcal{W} . The derivation of the derivative and Hessian of the objective function \mathcal{J}_3 given in Equation (4.47) follows similar

steps as in (K. Rahbar, J. Reilly and J. Manton, 2004), however we derive it for the non-square case.

$$\begin{aligned}
\mathcal{J}_3(\mathcal{W} + \delta \mathbf{Z}) &= \mathcal{J}_3(\mathcal{W}) + 4\delta \mathbb{R}\{tr(\mathbf{Z}^H off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}})\} \\
&+ 2\delta^2 tr(\mathbf{Z}^H off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}}) \\
&+ \mathbf{Z}^H off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \\
&+ 2\delta^2 \mathbb{R}\{tr(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H))\} \\
&+ O(\delta^3)
\end{aligned} \tag{A 4}$$

The last three terms of Equation (A.4) can be rewritten as,

$$\begin{aligned}
tr(\mathbf{Z}^H off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}}) \\
= vec(\mathbf{Z})^H (\mathbf{R}_{\mathcal{X}\mathcal{X}}^T \otimes off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) vec(\mathbf{Z})
\end{aligned} \tag{A 5}$$

$$\begin{aligned}
tr(\mathbf{Z}^H off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}}) \\
= vec(\mathbf{Z})^H vec(off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}}) \\
= vec(\mathbf{Z})^H (\mathbf{R}_{\mathcal{X}\mathcal{X}}^T \mathcal{W}^T \otimes \mathbf{I}_N) vec(off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) \\
= vec(\mathbf{Z})^H (\mathbf{R}_{\mathcal{X}\mathcal{X}}^T \mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{off} vec(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \\
= vec(\mathbf{Z})^H (\mathbf{R}_{\mathcal{X}\mathcal{X}}^T \mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{off} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^* \otimes \mathbf{I}_N) vec(\mathbf{Z})
\end{aligned} \tag{A 6}$$

$$\begin{aligned}
tr(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) \\
= vec(\mathbf{Z})^T \mathbf{P}_{vec}^{(N,N)} vec(\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) \\
= vec(\mathbf{Z})^T \mathbf{P}_{vec}^{(N,N)} (\mathbf{I}_N \otimes \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) vec(off(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) \\
= vec(\mathbf{Z})^T \mathbf{P}_{vec}^{(N,N)} (\mathbf{I}_N \otimes \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathbf{P}_{off} vec(\mathbf{Z} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \\
= vec(\mathbf{Z})^T \mathbf{P}_{vec}^{(N,N)} (\mathbf{I}_N \otimes \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathbf{P}_{off} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^* \otimes \mathbf{I}_N) vec(\mathbf{Z}) \\
= vec(\mathbf{Z})^T (\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H \otimes \mathbf{I}_N) \mathbf{P}_{vec}^{(N,N)} \mathbf{P}_{off} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^* \otimes \mathbf{I}_N) vec(\mathbf{Z})
\end{aligned} \tag{A 7}$$

After substituting the terms in Equations (A.5-A.7) into Equation (A.4), we equate the corresponding terms for the gradient and Hessian matrices with the second order Taylor series approximation in Equation (A.3) to obtain,

$$\begin{aligned}
\mathbf{G}_{\mathcal{W}} &= 2\{off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}}^H \\
&+ off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}}^H \mathcal{W}^H) \mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}}\},
\end{aligned} \tag{A 8}$$

and

$$\begin{aligned}
\{\mathbf{H}_{\mathcal{W}} + \mathbf{C}_{\mathcal{W}}\} &= 2\{(\mathbf{R}_{\mathcal{X}\mathcal{X}}^* \otimes off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H)) \\
&+ (\mathbf{R}_{\mathcal{X}\mathcal{X}}^T \otimes off(\mathcal{W} \mathbf{R}_{\mathcal{X}\mathcal{X}}^H \mathcal{W}^H)) \\
&+ (\mathbf{R}_{\mathcal{X}\mathcal{X}}^T \mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{off} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^* \otimes \mathbf{I}_N) \\
&+ (\mathbf{R}_{\mathcal{X}\mathcal{X}}^* \mathcal{W}^T \otimes \mathbf{I}_N) \mathbf{P}_{off} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^T \otimes \mathbf{I}_N) \\
&+ (\mathbf{R}_{\mathcal{X}\mathcal{X}} \mathcal{W}^H \otimes \mathbf{I}_N) \mathbf{P}_{vec}^{(N,N)} \mathbf{P}_{off} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^* \otimes \mathbf{I}_N) \\
&+ (\mathbf{R}_{\mathcal{X}\mathcal{X}}^H \mathcal{W}^H \otimes \mathbf{I}_N) \mathbf{P}_{off} \mathbf{P}_{vec}^{(N,N)} (\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X}}^T \otimes \mathbf{I}_N)\}
\end{aligned} \tag{A 9}$$

Considering all $k = 1, 2, \dots, K$ time window frames, over all time lags $\tau = -\tau_{min}, \dots, \tau_{max}$ these expressions are expanded to \mathbf{G}_3 and \mathbf{H}_3 given in Table 4.1. The constraint function is given as

$$\begin{aligned}
\mathcal{J}_4(\mathcal{W}) &= \|ddiag(\mathcal{W} \mathcal{W}^H - \mathbf{I}_N)\|_F^2 \\
&= tr\{ddiag(\mathcal{W} \mathcal{W}^H - \mathbf{I}_N) ddiag(\mathcal{W} \mathcal{W}^H - \mathbf{I}_N)\} \\
&= tr\{ddiag(\mathcal{W} \mathcal{W}^H) ddiag(\mathcal{W} \mathcal{W}^H) \\
&- 2ddiag(\mathcal{W} \mathcal{W}^H) + \mathbf{I}_N\}
\end{aligned} \tag{A 10}$$

Using the same approach as in Equation (A.4), we derive,

$$\begin{aligned}
\mathcal{J}_4(\mathcal{W} + \delta \mathbf{Z}) &= \mathcal{J}_4(\mathcal{W}) + 4\delta \mathbb{R}\{tr(\mathbf{Z}^H (ddiag(\mathcal{W} \mathcal{W}^H) - \mathbf{I}_N) \mathcal{W})\} \\
&+ 2\delta^2 \{vec(\mathbf{Z})^H vec((ddiag(\mathcal{W} \mathcal{W}^H) - \mathbf{I}_N) \mathbf{Z}) \\
&+ vec(\mathbf{Z})^H vec(ddiag(\mathbf{Z} \mathcal{W}^H) \mathcal{W})\} \\
&+ \delta^2 \{vec(\mathbf{Z})^T vec(ddiag(\mathcal{W} \mathbf{Z}^H) \mathcal{W}) \\
&+ vec(\mathbf{Z})^T \mathbf{P}_{vec}^{(N,M,Q)} vec(\mathcal{W}^H ddiag(\mathbf{Z} \mathcal{W}^H))\} \\
&+ O(\delta^3)
\end{aligned} \tag{A 11}$$

The last four terms of Equation (A.11) can be rewritten as

$$\begin{aligned} \text{vec}((d\text{diag}(\mathcal{W}\mathcal{W}^H) - \mathbf{I}_N)\mathbf{Z}) \\ = (\mathbf{I}_{MQ} \otimes d\text{diag}(\mathcal{W}\mathcal{W}^H - \mathbf{I}_N))\text{vec}(\mathbf{Z}) \end{aligned} \quad (\text{A } 12)$$

$$\begin{aligned} \text{vec}(\text{diag}(\mathbf{Z}\mathcal{W}^H)\mathcal{W}) \\ = (\mathcal{W}^T \otimes \mathbf{I}_N)\text{vec}(d\text{diag}(\mathbf{Z}\mathcal{W}^H)) \\ = (\mathcal{W}^T \otimes \mathbf{I}_N)\mathbf{P}_{\text{diag}}\text{vec}(\mathbf{Z}\mathcal{W}^H) \\ = (\mathcal{W}^T \otimes \mathbf{I}_N)\mathbf{P}_{\text{diag}}(\mathcal{W}^* \otimes \mathbf{I}_N)\text{vec}(\mathbf{Z}) \end{aligned} \quad (\text{A } 13)$$

$$\begin{aligned} \text{vec}(d\text{diag}(\mathcal{W}\mathbf{Z}^H)\mathcal{W}) \\ = (\mathcal{W}^H \otimes \mathbf{I}_N)\text{vec}(d\text{diag}(\mathcal{W}\mathbf{Z}^H)) \\ = (\mathcal{W}^H \otimes \mathbf{I}_N)\mathbf{P}_{\text{diag}}\text{vec}(\mathcal{W}\mathbf{Z}^H) \\ = (\mathcal{W}^H \otimes \mathbf{I}_N)\mathbf{P}_{\text{diag}}(\mathbf{I}_N \otimes \mathcal{W}^*)\mathbf{P}_{\text{vec}}^{(N \ MQ)T}\text{vec}(\mathbf{Z}) \end{aligned} \quad (\text{A } 14)$$

$$\begin{aligned} \mathbf{P}_{\text{vec}}^{(N \ MQ)}\text{vec}(\mathcal{W}^H d\text{diag}(\mathbf{Z}\mathcal{W}^H)) \\ = \mathbf{P}_{\text{vec}}^{(N \ MQ)}(\mathbf{I}_N \otimes \mathcal{W}^H)\text{vec}(d\text{diag}(\mathbf{Z}\mathcal{W}^H)) \\ = \mathbf{P}_{\text{vec}}^{(N \ MQ)}(\mathbf{I}_N \otimes \mathcal{W}^H)\mathbf{P}_{\text{diag}}\text{vec}(\mathbf{Z}\mathcal{W}^H) \\ = \mathbf{P}_{\text{vec}}^{(N \ MQ)}(\mathbf{I}_N \otimes \mathcal{W}^H)\mathbf{P}_{\text{diag}}(\mathcal{W}^* \otimes \mathbf{I}_N)\text{vec}(\mathbf{Z}) \end{aligned} \quad (\text{A } 15)$$

After substituting the terms in Equations (A.12-A.15) into Equation (A.11), we equate the corresponding terms for the Jacobian and constraint Hessian matrices with the second order Taylor series approximation in Equation (A.3) to obtain,

$$\mathbf{G}_{\mathcal{W}} = 4(d\text{diag}(\mathcal{W}\mathcal{W}^H) - \mathbf{I}_N)\mathcal{W} \quad (\text{A } 16)$$

and

$$\begin{aligned} \{\mathbf{H}_{\mathcal{W}} + \mathbf{C}_{\mathcal{W}}\} &= 4\{(\mathbf{I}_{MQ} \otimes d\text{diag}(\mathcal{W}\mathcal{W}^H - \mathbf{I}_N)) \\ &+ (\mathcal{W}^T \otimes \mathbf{I}_N)\mathbf{P}_{\text{diag}}(\mathcal{W}^* \otimes \mathbf{I}_N)\} \\ &+ 2\{(\mathcal{W}^H \otimes \mathbf{I}_N)\mathbf{P}_{\text{diag}}(\mathbf{I}_N \otimes \mathcal{W}^*)\mathbf{P}_{\text{vec}}^{(N \ MQ)T} \\ &+ \mathbf{P}_{\text{vec}}^{(N \ MQ)}(\mathbf{I}_N \otimes \mathcal{W}^H)\mathbf{P}_{\text{diag}}(\mathcal{W}^* \otimes \mathbf{I}_N)\} \end{aligned} \quad (\text{A } 17)$$

The Jacobian and constraint Hessian matrices, \mathbf{G}_4 and \mathbf{H}_4 , are given in Table 4.1.