Blind Separation of Heavy Tail Signals

Methods for Blind Separation of Linear Mixtures of Signals with Heavy Tail Distributions

Preben Kidmose

LYNGBY 2001

IMM-PHD-2001-88

IMM

Blind Separation of Heavy Tail Signals

Methods for Blind Separation of Linear Mixtures of Signals with Heavy Tail Distributions

Preben Kidmose

LYNGBY 2001

IMM-PHD-2001-88

TECHNICAL UNIVERSITY OF DENMARK
Informatics and Mathematical Modelling
Richard Petersens Plads, Building 321
DK-2800 Kongens Lyngby, Denmark

Ph.D. Thesis
In partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Title: Blind Separation of Heavy Tail Signals
Methods for Blind Separation of Linear Mixtures of Signals with Heavy Tail Distributions

Author: Preben Kidmose

TECHNICAL UNIVERSITY OF DENMARK
Informatics and Mathematical Modelling
Richard Petersens Plads, Building 321
DK-2800 Kongens Lyngby, Denmark
IMM-PHD-2001-88

Lyngby 2001-09-01

Printed by IMM, Technical University of Denmark

ISSN 0909–3192

PREFACE

The present thesis has been submitted in partial fulfillment of the requirements for the Ph.D. degree in electrical engineering. The work documented in this thesis was initiated in September 1998 and completed in September 2001 at Informatics and Mathematical Modelling[1] (IMM), Technical University of Denmark (DTU).

The thesis is on method for blind separation of linear mixtures of signals with heavy tail distributions. Blind signal separation has become an important part of statistical signal processing, and this thesis is a contribution in the constitution of blind signal separation. The establishment of blind signal separation methods provides new signal processing capabilities that would not be possible otherwise. Blind signal separation is succesfully applied in diverse fields, and in this work the foucs is on applications in audio processing.

The conditions under which the Ph.D. study has been carried out have from time to time been quite turbulent, and it is appropriate to give that a few remarks. The Ph.D. project was established on the initiative of John Aasted Sørensen who at that time was Associate Professor at Section for Digital Signal Processing, Department of Mathematical Modelling. Initially John Aasted Sørensen was supervisor, and Associate Professor Steffen Duus Hansen and Professor Per Christian Hansen were co-supervisors. Unfortunately Steffen had blood cancer and passed away August 2nd 1999. John Aasted Sørensen left the department on November 1st 2000, and Professor Lars Kai Hansen became my new supervisor. Per Christian Hansen, unfortunately, never really became involved in the project, because it took a different direction than first assumed.

Acknowledgements

First I want to acknowledge all the supervisors who have been involved in the project. John Aasted Sørensen who initiated the project and arranged all the funding and practicalities. Steffen Duus Hansen for the way he was as a human being. And Lars Kai Hansen who was willing to take the responsibility for the project when everything nearly collapsed. I appreciate the effort Lars Kai Hansen put into the project, and for his inspiring comments on the work that I had already done.

I wish to acknowledge all the colleagues in the DSP-group at IMM for providing an inspiring and pleasant environment. Special thanks to Peter Søren Kirk Hansen, my office-mate for three years, for nice company. Thanks to Jørgen Madsen, Dorthe Kidmose and Thomas Fabricius for helping with the proofreading of the manuscript.

From March to September 2000 I visited the Signal and Image Processing Institute (SIPI) at University of Southern California (USC), USA. Thanks to Assistant Professor Chris Kyriakakis for inviting me

[1]The Ph.D. project was initiated in the Section for Digital Signal Processing, Department of Mathematical Modelling, Technical University of Denmark. Due to a reorganization of the department structure at the university, and a following reorganization of the department, the project was completed at Informatics and Mathematical Modelling, Technical University of Denmark

to USC. Thanks to Ph.D. Student Panayiotis Georgiou, my office-mate at SIPI, for his company in the office and for helping me with a lot of practical things.

I acknowledge DTU and the Danish Research Councils for the financial support for the Ph.D. study. The Otto Mønsteds Foundation is acknowledged for financial support for the research visit at SIPI, USC. The Danish Research Councils through the THOR Center for Neuroinformatics is acknowledged for partially financing the trips to ICASSP 2001 and NSIP 2001.

Finally, I am indebted to my wife Hanne for her endless patience, tolerance and support throughout this project.

Publication note

Parts of the research presented in this thesis have previously been published in conference proceedings, and the following papers have been presented at conferences:

- Preben Kidmose, Adaptive Filtering for Non-Gaussian Processes, *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pp. 424-427, 2000.
- Preben Kidmose, Alpha-Stable Distributions in Signal Processing of Audio Signals, *Proceedings of 41st Conference on Simulation and Modelling, Scandinavian Simulation Society*, pp. 87-94, 2000.
- Thomas Fabricius, Preben Kidmose and Lars Kai Hansen, Dynamic Components of Linear Stable Mixtures from Fractional Low Order Moments, *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pp. 3957 -3960, 2001.
- Preben Kidmose, Independent Components Analysis Using the Spectral Measure for Alpha-Stable Distributions, *Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, pp. 1-5, 2001.

The thesis is printed by IMM, Technical University of Denmark. The thesis is avaiable electronically from `http://www.imm.dtu.dk/documents/ftp/publphd.html`.

Technical University of Denmark, September 2001

Preben Kidmose

ABSTRACT

Blind Separation of Heavy Tail Signals

This Ph.D. thesis is on methods for blind signal separation of signals with heavy tail distributions. The thesis is partly motivated by applications in digital signal processing of audio signals, and partly by the theoretical basis for blind separation of heavy tail signals. Initially an introduction to blind signal separation and the inherent problems are given, the signal models are introduced, and the models are discussed in relation to different applications.

The thesis introduces different heavy tailed distributions that have potential application in the field of signal processing, and special emphasis is on a certain class of distributions - the stable distributions. The class of stable distributions have some attractive analytical characteristics and it turns out that they are suitable for modelling many physical signals. Many of the fundamental assumptions in classical statistical signal processing does not apply to heavy tail signals in general. The thesis establishes the necessary theoretical basis for statistical signal processing and for blind separation of heavy tail signals.

The thesis introduces the most important existing methods for blind signal separation and the methods are discussed in relation to signals with heavy tail distributions. One of the most important methods for blind signal separation is based on the maximum likelihood principle. It turns out that blind signal separation methods based on the maximum likelihood method is appropriate to deal with heavy tail signals, that the method is relatively robust to deviations in the model of the distributions of the signals, and that the methods work even for stable distributions. A class of methods for blind signal separation is based on joint diagonalization of matrices; the thesis propose some variants of these methods and a unifying presentation of these methods are given. Finally a new method for blind signal separation of heavy tail signals, that is based on the spectral measure for stable distributions is proposed. The method has two remarkable properties, the number of independent signals and the corresponding basis vectors in a general rectangular mixing matrix can be identified.

The thesis focuses on applications in processing of audio signals, and an empirical examination of a broad class of audio signals shows that the class of stable distributions are suitable for modelling audio signals. The methods for blind signal separation are evaluated on various audio signals from synthetic mixtures and from signals from an experimental teleconference system.

RESUMÉ (IN DANISH)

Blind separation af signaler med tunghale fordelinger

Denne ph.d. afhandling omhandler metoder til blind signalseparation af signaler med tunghale fordelinger. Afhandlingen er dels motiveret af en række anvendelser indenfor digital signalbehandling af audiosignaler, og dels af det teoretisk fundament for blind separation af tunghale signaler. Der indledes med en introduktion til blind signalseparation og de dertilhørende grundliggende problemstillinger; og der indføres forskellige signalmodeller og modellerne diskuteres i forhold til forskellige anvendelser.

Afhandlingen introducerer forskellige tunghale fordelinger der har vist sig at have potentiel anvendelse indenfor signalbehandling, og der fokuseres specielt på en bestemt klasse af fordelinger - de stabile fordelinger. Klassen af stabile fordelinger har en række gode analytiske egenskaber og samtidig viser det sig at familien af stabile fordelinger er velegnet til at modellere mange fysiske signaler. Mange af de grundlæggende antagelser i klassisk statistisk signalbehandling er i det generelle tilfælde ikke gyldige for signaler med tunghale fordelinger. Afhandlingen etablerer det nødvendige teoretiske fundament for statistisk signalanalyse og for blind signalseparation af signaler med tunghale fordelinger.

Afhandlingen introducerer de vigtigste eksisterende metoder til blind signalseparation og metoderne diskuteres i forhold til signaler med tunghale fordelinger. En af de vigtigste metoder til blind signalseparation er baseret på maksimum likelihood princippet. Det viser sig at blind signalseparation baseret på maksimum likelihood metoden er velegnet til at håndtere tunghale fordelinger, at metoden er relativt robust overfor afvigelser i modellen for fordelingen af signalerne, og at metoden fungerer selv for stabile fordelinger. En klasse af metoder til blind signalseparation er baseret på diagonalisering af matricer; afhandlingen introducerer nogle varianter af disse metoder og en samlet fremstilling præsenteres. Endelig introduceres en ny metode til blind signalseparation af tunghale signaler, der er baseret på det spektrale mål for stabile fordelinger. Metoden har to bemærkelsesværdige egenskaber, antallet af komponenter og de tilhørende basis vektorer i en generel rektangulær blandingsmatrix (instantaneous mixture) kan identificeres.

I afhandlingen fokuseres der på anvendelser indenfor signalbehandling af audiosignaler, og der gennemføres en analyse der viser at en stor klasse af audiosignaler kan modelleres med stabile fordelinger. Metoderne til blind signalseparation evalueres på forskellige audiosignaler, både med signaler der er syntetisk blandet og på realistiske signaler fra et eksperimentielt telekonference system.

CONTENTS

CHAPTER 1

INTRODUCTION

Blind signal separation is a fundamental problem in signal processing and it has received considerable attention throughout the last decade. The reason why blind signal separation has received this attention is because of a vast number of potential applications and because inherent in the blind signal separation problem there are a number of interesting theoretical problems. Even though blind signal separation has now established itself in the literature [41, 62], there are still many unsolved problems and it is indeed still an active field of research.

Blind signal separation addresses the problem, from an observable mixture of source signals, to separate or ideally reconstruct the unknown source signals. The term *blind* refers to the fact that the source signals and the way the source signals are mixed is unknown. The mixtures of the source signals are termed the observable signals, and the model of the mixing of the source signals is referred to as the mixing system. The separated signals are obtained from the observable signals by means of a separation system; in Figure 1.1 the signal model is depicted as a block diagram.

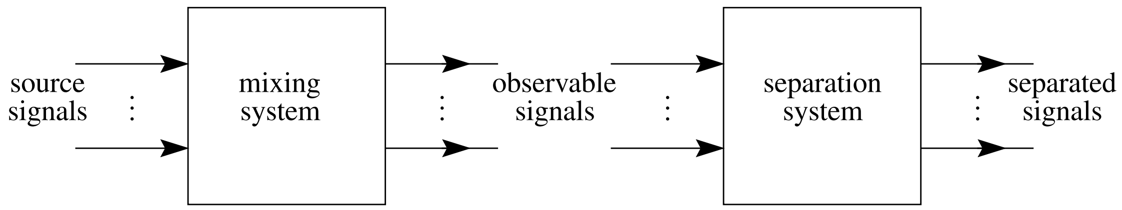


Figure 1.1 Signal model for the blind signal separation problem.

The statistical principles for blind signal separation has evolved from the field of classical statistical signal processing and from the field of Independent Component Analysis (ICA), and with significant contributions from the field of information theory and artificial neural networks.

As the title of the thesis suggests, the emphasis in this work is on a class of signals that has *heavy tails*. Heavy tail signals are likely to exhibit large observation and has often an impulsive nature – and it turns out that a broad class of audio signals have heavy tail nature. The term *heavy tail* refers to that the probability density functions of the signals have relative large mass in the *tails*.

This introductory chapter provides a motivation for the thesis and a brief discussion of the fundamental problems in blind signal separation. The thesis is partly motivated by applications in audio signal processing and partly by the theoretical basis for blind signal separation of heavy tail signals. To put the thesis into a larger perspective some of the areas where independent component analysis and blind signal separation have found application are mentioned in Section 1.1. The thesis put emphasis on audio applications, and a typical application for blind separation of audio signals – a teleconference system –

is considered. In Section 1.2 the general blind signal separation problem and the inherent limitations are discussed. The scope of the thesis is formed by Section 1.1 and 1.2. Finally, in Section 1.3, the content and outline of the thesis is described.

1.1 Applications of Blind Signal Separation

On the application side this thesis put emphasis on blind signal separation of audio signals. However, it is appropriate before proceeding to mention some of the other fields where independent component analysis and blind signal processing have found applications. Furthermore, it should be noticed that the main part of the thesis does not have any specific application in mind, and the application in audio signal processing is mainly for the purpose of relating the theory to practical applications.

Blind signal separation techniques have found wide use in the field of biomedical signal processing; early pioneering works is found in [91, 92] where ICA is used for analysis of MEG and EEG signals, and in [68] where ICA is used for analysis of activity patterns in functional MRI data. In image processing [43] applied an ICA method proposed in [69] on short image sequences, and an equivalent method is used much earlier in [87] where the temporal correlation structure of multi-spectral images is explored. Another field where ICA has found applications is in multimedia signal processing [44], and early work on textual data analysis is found in [49, 58]. An early and pioneering work in the field of array signal processing is found in [24]. ICA is also merging into areas as communications and analysis of financial time series.

For speech and audio signals blind signal separation has potential in applications where there is a need for noise reduction, speech enhancement, and for separation of multiple sources into multiple audio tracks. Not only will the development and implementation of blind signal separation techniques improve the audio quality and user-friendliness in applications such as teleconference systems, mobile phones, hearing aids, voice controlled devices, etc., but it may also open up for new applications in these areas.

1.1.1 Blind Signal Separation in Audio Applications

The extension from single microphone to multiple microphone systems combined with the proliferation of integrated media technologies and the steady increase in computing power make a fundamental change in the possibilities in audio and multimedia applications. One of the possibilities from multiple microphone signals is to extract or separate individual speakers and sound sources. A scenario for this application is in a teleconference system where it may be desirable to avoid personal/wearable microphones. Recently, considerable research has been carried out for blind signal separation in the teleconference application [48, 64, 83, 61, 85], and in the following this scenario is discussed in more detail. However, it is interesting before proceeding to notice some related methods and advances in the field of audio applications and blind source separation, which are beyond the scope of this thesis. In [81] a method for separation of speech signals by a non-stationary reweigthing of frequency sub-bands from a one microphone signal is proposed. In [79, 59] ICA is used to feature extraction in speech signals with application in automatic speech recognition. Also a vast amount of research is going on with regards to integration/fusion of video and audio signals in speech recognition, source localization, etc..

Integrated media technologies and multiple microphone systems combined with the increasing computing power and high-bandwidth networks establish the technological basis for high-quality teleconference applications. In a teleconference system the goal is to create a virtual conference environment for multiple participants located at different sites – and maybe even create the illusion of proximity among multiple participants. In Figure 1.2[1] an example of a multiple-site teleconference environment with mul-

[1]The figure is partly a reproduction from [38] with permission from P. G. Georgiou.

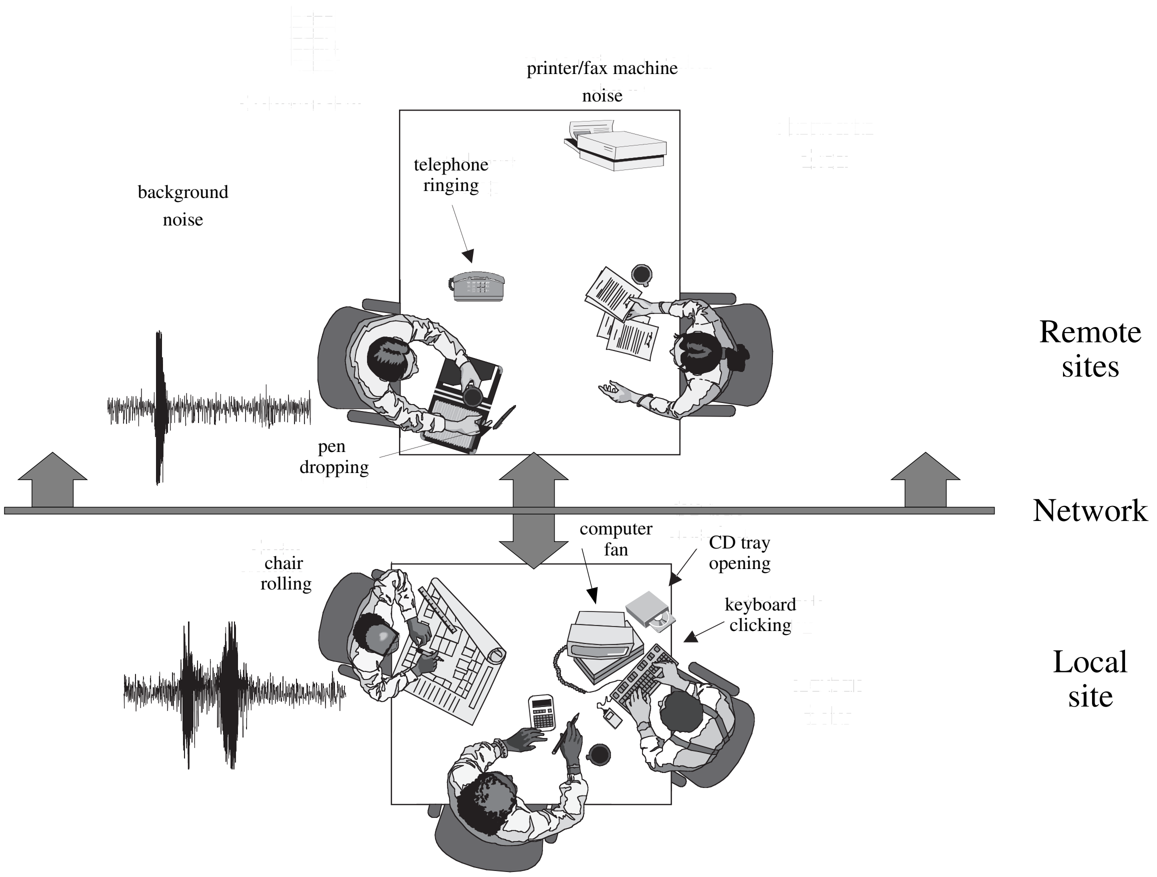


Figure 1.2 A typical teleconference environment. The aural scene is set by the speaking participants, various kinds of background noise sources, and the acoustics of the room.

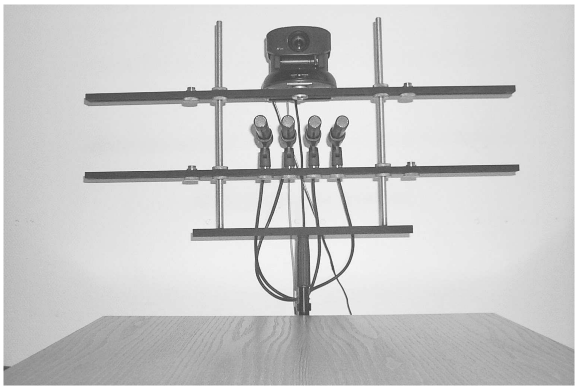
tiple participants at different sites is depicted. The *aural scene* is set by the speaking participants, various kinds of background noise sources, the acoustics of the room, etc. To create a convenient and realistic aural scene at a local site it is necessary to integrate the important aural cues from the remote sites into the local aural scene in a credible way.

The frontend of the teleconference system must provide an input/output audio interface. The task of the audio interface is to integrate the audio streams from the remote sites into an augmented virtual conference environment at the local site; and to capture the aural cues from the local site and stream the individual audio streams to the remote sites. The idea is to use blind source separation techniques to capture the aural cues, speech signals as well as other important aural cues, as separate audio streams; and this is the application that is in mind in the practical part of this thesis.

Audio signals from a realistic teleconference environment has been available from the RTMM experimental teleconference system[2]. The sites of the RTMM system are equipped with microphone arrays and provide a high-quality audio interface. One of the sites of the experimental teleconference system is shown in Figure 1.3. For technical specification of the audio hardware and audio interface see [45].

[2]The Real Time Multi-Modal (RTMM) project is a research project in collaboration between IMM and COM at the Technical University of Denmark and the Department of Information and Media Science at the University of Århus. The RTMM project is financially supported by The Danish Research Council, and Danish Centre for Information Technology (CIT).

(a)



(b)

Figure 1.3 The RTMM experimental teleconference setup. (a) One of the teleconference sites equipped with a microphone-array, a camera and loudspeakers. (b) The microphone stand with a 4-microphone array and a camera.

1.2 The Blind Signal Separation Problem

The aim of this section is to give an introduction to the fundamental issues of blind signal separation. A discussion of the generative model and an intuitive introduction to the blind signal separation problem is provided, and it is intended to give an easy accessible introduction.

1.2.1 The Generative Model

The generative model reflects the model that is believed to generate the observable signals, i.e. it describes the mixture system and a probabilistic model of the source signals.

Generally, the observed signals are assumed to be a linear combination of independent source signals, i.e. the mixture is either an instantaneous mixture or a convolutive mixture. The choice of generative model is strongly application dependent. For narrow band signals, or signals where there are no temporal combinations, the instantaneous mixture is suitable. The more general convolutive mixture is appropriate for modelling multi-path mixtures, e.g. time delayed reflections.

In the literature the instantaneous mixture has received most attention, and all the ICA methods are based on the instantaneous mixture model. Most of the fundamental problems in blind signal separation can be discussed based on the real instantaneous mixture, and the discussions in this section are on the basis of that model whenever it suffices.

A general knowledge from mathematical modelling is that strong a priori knowledge in general results in effective models of the cost of robustness against deviations from the a priori assumptions. This is indeed also the case for the generative model in the blind signal separation problem, thus the choice of generative model has crucial influence on the performance of the blind signal separation.

Instantaneous Mixture Let $s_{n,k}$ denote the nth source signal and let $x_{m,k}$ denote the mth observable signal, both with time index k. The term *instantaneous mixture* refers to the fact that there is no time-shift or convolution involved in the transfer from the source signals to the observable signals. Let a_{mn} denote the coefficient from source signal n to observable signal m, the mth observable signal can then be written as the sum

$$x_{m,k} = \sum_{n=1}^{N} a_{mn} \; s_{n,k}$$

It is convenient to formulate the instantaneous mixture on a vector form. By arranging the signals in vectors, $\mathbf{s}_k = [s_{1,k}, \ \ldots \ , s_{N,k}]^T$ and $\mathbf{x}_k = [x_{1,k}, \ \ldots \ , x_{M,k}]^T$, the instantaneous mixture can be written as

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is termed the *mixture matrix*. The number of source signals, N, and the number of observable signals, M, determine the dimension of the mixture. The case $N < M$ (the number of source signals < the number of observable signals) is referred to as the over-determined case; and the case $N > M$ are referred to as the under-determined case.

Convolutive Mixture The convolutive mixture is a model where each of the observable signals is a sum of filtered source signals. As for the instantaneous mixture, let $s_{n,k}$ denote the nth source signal, and $x_{m,k}$ denote the mth observable signal. Furthermore, let $A_{mn,l}$ denote the lth filter coefficient in the filter from the nth source signal to the mth observable signal. The mth observable signal can then be expressed as a sum of convolutions

$$x_{m,k} = \sum_{n=1}^{N} \sum_{l=0}^{L-1} A_{mn,l} \; s_{n,k-l}$$

As for the instantaneous mixture it is convenient to arrange the signals into vectors. Let $\mathbf{s}_k$ denote the source signal vector and $\mathbf{x}_k$ the observable signal vector; the convolutive mixture can then be written as

$$\mathbf{x}_k = \sum_{l=0}^{L-1} \mathbf{A}_l \; \mathbf{s}_{k-l}$$

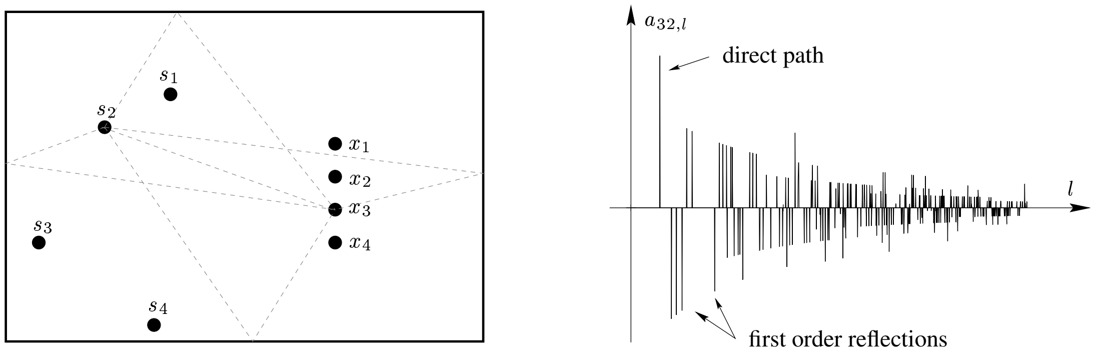
where $\mathbf{A}_l \in \mathbb{R}^{M \times N}$ for $l = 0, 1, \ldots, L-1$ is matrices of filter coefficients, and $A_{mn,l}$ is the (m, n)th element of $\mathbf{A}_l$. Another and even more compact notation is obtained in the z-domain; the z-domain mixture matrix $\mathbf{A}(z)$ is the polynomial matrix

$$\mathbf{A}(z) = \sum_{l=0}^{L-1} \mathbf{A}_l \; z^{-l}$$

The (m, n)th polynomium in the mixture matrix is the transfer function from the nth source signal to the mth observable signal. The convolutive mixture model can then be written as $\mathbf{x}(z) = \mathbf{A}(z) \; \mathbf{s}(z)$.

A practical example where the convolutive mixture model is appropriate, is in the modelling of the mixture of acoustical signals in a reverberant room. Due to reflections from the walls of the room the impulse response between two points in the room describes a multi-path transfer function. In Figure 1.4(a) this is illustrated in a room with 4 sources and 4 microphones, and in Figure 1.4(b) a typical acoustical impulse response is depicted. The model of an acoustical mixture as depicted in Figure 1.4 is in fact an appropriate mixture model for the multiple microphone teleconference system depicted in Figure 1.3.

In general, the blind signal separation problem for convolutive mixtures is a considerable harder problem to solve than for instantaneous mixtures. However both the instantaneous and convolutive mixture models are interesting from a theoretical as well as an application point of view, and both models are treated in this thesis.



(a) (b)

Figure 1.4 The impulse response of a room. (a) Depiction of a room with 4 sources ($s_1 - s_4$) and 4 microphones ($x_1 - x_4$). The direct path and the 4 first order reflections from source s_2 to microphone x_3 are depicted with dashed lines. (b) A typical example of an impulse response, the direct path and the first 4 order reflections are easily recognized and are indicated with arrows. The impulse response is synthetically generated using the image method [4, 14].

Probabilistic Source Signal Model As previously mentioned the generative model describes the mixture of the source signals and the probabilistic structure of the individual source signals. An adequate probabilistic description of the source signals consists of the probability density functions and a description of the temporal structure of the sources.

Specific applications can justify assumptions made on the source signals probability density functions; and signals are generally assumed to have either a continuous or discrete density function. Typical signals with discrete density functions are communication signals or other finite alphabet signals like text sequences. Acoustical signals or biomedical signals are typical examples of signals with continuous density functions. In this thesis only continuously distributed signals are considered and special attention is paid to signals with heavy tail density functions.

The fundamental assumption in all blind source separation methods is that the source signals are mutually independent, and furthermore independent component analysis methods normally assumes that the source signals are temporally independent. For more general blind source separation methods the temporal structure of the source signals are typically modelled as linear stochastic processes. Thus, the generative model can model source signals with or without temporal structure and both models are interesting from a theoretical as well as from a practical point of view.

1.2.2 An Intuitive Introduction to the Blind Source Separation Problem

The objective of the blind source separation is, based on the observable signals, to recover the source signals. A useful example for an intuitive discussion of the blind source separation problem is the mixture of acoustical sources as depicted in Figure 1.4(a), where the transfer function from each source to each microphone is a multi-path impulse response as depicted in Figure 1.4(b). Each of the microphone signals is thus a convolutive mixture of the 4 source signals.

It is instructive first to consider the blind signal separation for the case where the generative model is an instantaneous mixture. Let the mixture matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $M = N$ and let $\text{rank}(\mathbf{A}) = M$. Since $\mathbf{A}$ has an inverse, the source signals can be recovered from the observable signals as $\mathbf{s}_k = \mathbf{A}^{-1}\mathbf{x}_k$. It is useful to introduce the separating matrix $\mathbf{D}$ and a vector of separated signals $\mathbf{u} = \mathbf{D}\mathbf{x}$. It appears that if the separating system is the inverse of the mixture system, then the source signals is completely separated. Notice that inverse systems only exists for $M = N$. For over-determined instantaneous mixture, $N < M$, the separating system can be obtained as the pseudo-inverse of the mixture matrix, and for the under-determined mixture there is no non-trivial solution to the separating system.

For the convolutive mixture model the separating system that recovers the source signals is the inverse of the convolutive mixture. In the examination of the separating system it is useful to consider the z-transformed of the convolutive mixture, $\mathbf{A}(z)$, and for non-singular mixtures the inverse of the convolutive mixture is $\mathbf{A}^{-1}(z) = \operatorname{adj}\mathbf{A}(z)/\det\mathbf{A}(z)$. The adjugate matrix, $\operatorname{adj}\mathbf{A}(z)$, is a polynomial matrix; and the determinant of the polynomial matrix is, except for unimodal polynomial matrices[3], a polynomium in z. Since the determinant is a common factor for all the observable signals it appears that the adjugate matrix represents a separating system. In some applications there is a need to distinguish between a separating system, $\mathbf{D}(z) = \operatorname{adj}\mathbf{A}(z)$, and a recovering system, $\mathbf{D}(z) = \mathbf{A}^{-1}(z)$. For FIR mixtures the separating system is also a FIR system, whereas the recovering system in general is an IIR system. For acoustical systems it is interesting to notice that the acoustical impulse response in general is a non-minimum phase system and thus the impulse response of a single acoustical channel is in general not invertible [70]. However, for multiple channel systems a separating system exists even though the individual impulse responses are non-minimum phase.

From the previous it appears that a separating system is closely related to the inverse of the mixture system. Even though this is a reasonable conclusion it does not give the solution for the blind signal separation problem. However, it is possible to get an intuitive and abstract insight in the nature of the problem without too many technicalities. For that purpose it is convenient to introduce an abstract concept of measuring information. Consider an information measure $I_{\mathbf{u}}$ that measures the information between the components in the vector of separated signals $\mathbf{u}$. Let $I_{\mathbf{u}}$ be a positive semidefinite function of $\mathbf{u}$ where $I_{\mathbf{u}} = 0$ if and only if the elements in $\mathbf{u}$ do not contain mutual information. The blind source separation problem is then to minimize $I_{\mathbf{u}}$ with respect to the separating system $\mathbf{D}(z)$.

Consequences of the Blind Approach If the probability distributions of the source signals are identical and no other a priori knowledge of the sources are available, then the blind signal separation has a permutation and scaling ambiguity. For the instantaneous mixture model the scaling and permutation ambiguity means that any two rows of the separating matrix can be interchanged and scaled arbitrarily.

The separating system corresponding to a convolutive mixture also has a permutation and scaling ambiguity. However, the scaling ambiguity has even further consequences for the separating system in the case of convolutive mixtures. It appears that if no a priori knowledge is available of the temporal structure of the sources, then the separation system has a convolutive ambiguity. Thus, the separating system is blind to whether temporal structure origins from the sources or from the mixture.

These consequences for blind signal separations are readily obtained from the concept of measuring information between the separated signals.

A Spatial Interpretation of the Blind Separating System Related to blind signal separation are array signal processing and beamforming[4]. Beamforming is a spatiotemporal filtering technique that is based on the sensor array aperture, and the spatial response is referred to as the *beam*.

Blind signal separation methods do not rely on the sensor array aperture, however, if the array aperture is known then a spatial interpretation of the separation system is possible. This interpretation provides an intuitive insight in the separating system and links the field of blind signal separation and the field of beamforming.

A practical example of a spatial interpretation of a separating system is depicted in Figure 1.5. The microphone array is placed in the center of the polar coordinate system, and the polar plots depict the spatial responses of two of the channels in the separating system. A couple of interesting observations

[3]See [52] and references herein for the theory of polynomial matrices.

[4]See [50] for a classical text book on array signal processing

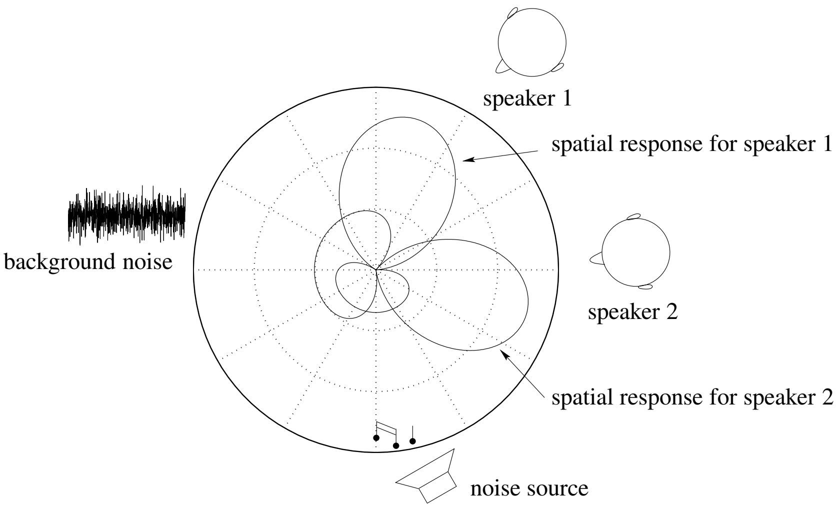


Figure 1.5 Spatial interpretation of a separating system. The graphs in the polar coordinate system depicts the amplitude of the spatial response for the channels of the separating system. Notice that the spatial response for speaker 1 is zero in the direction of speaker 2 and vise versa.

appear from the figure. The two spatial responses correspond to the channels of the separating system that separate speaker 1 and speaker 2. The spatial response corresponding to the channel with speaker 1 place a zero in the spatial response in the direction of speaker 2 and vise versa, and the direction of the source signal is in general not equal to the direction in which the spatial response has a maximum. The spatial response for speaker 1 and 2 has low gain in the direction of the background sources.

1.3 Content and Outline

In this section an overview of the content and organization of the thesis is provided.

Chapter 2 revives the fundamentals of statistical signal processing with higher order statistics and non-linear functions. Moreover the chapter establishes the probability and information theory necessary for the introduction to heavy tail distributions and the later development of blind signal separation methods.

Chapter 3 provides an introduction to the theory for heavy tail distributions with emphasis on the class of stable distributions. Some of the fundamental assumptions for classical statistical signal processing and blind signal separation do not apply to signals with heavy tail distributions, and this chapter establishes the basis for the later development of blind signal separation for heavy tail signals.

Chapter 4 reviews some of the most important methods for blind signal separation, and special attention is paid to the implication of heavy tail signals. Firstly, blind signal separation based on the maximum likelihood principle is considered, then methods based on joint diagonalization, and finally a novel method based on the spectral measure for stable distributions is introduced. Simulation examples based on synthetic signals exhibit the properties of the methods.

Chapter 5 focuses on the applications of the methods developed in Chapter 2, 3 and 4 in audio signal processing. It is demonstrated that heavy tail distributions are suitable for modelling a broad class of audio signals, and a new robust adaptive filtering method is proposed. Several examples demonstrate the applicability of blind signal separation methods in processing of audio signals, and a blind signal separation method is applied to signals from an experimental teleconference system.

At the end of each chapter there is a brief summary in which specific research contributions are also mentioned. Chapter 6 summarizes the work, concludes the thesis, and draws the lines for further research.

The papers published in relation to the Ph.D. study are included in Appendix A. At the end of the thesis a nomenclature, a bibliography, and an index are included. Special symbols, notations and abbreviations are introduced in the text the first time they appear, and the nomenclature is included for readers that do not read the complete thesis. All references in the thesis are linked to the bibliography, and a reference will appear as $[n]$, where n is an index to the bibliography.

CHAPTER 2

FUNDAMENTALS OF STATISTICAL SIGNAL PROCESSING

Classic statistical signal processing is dominated by methods based on second order statistics and assumptions on Gaussian signals. Because the second order statistics are sufficient for Gaussian signals, the assumption on Gaussian signals and methods based on second order statistics forms a strong pair. However, it is well known that second order statistics have fundamental and inherent limitations and that real world signals are rarely Gaussian. Furthermore the limitations of second order statistics are often in the way of the solution of the blind signal separation problem. To overcome these limitations this chapter extends the theory to higher order statistics and non-linear functions.

The objective of this chapter is to review the probability and information theory required for the development of the blind signal separation techniques in Chapter 4, and to establish terminology and notation. The chapter is intended to be an intuitive and accessible introduction and therefore a rigorous presentation is avoided. Furthermore a lot of small examples are given to illustrate the theory and to give intuitive insight into the blind signal separation problem.

The chapter is organized as follows. Section 2.1 introduces elementary probability functions and the theory of moment and cumulant functions. Section 2.2 reviews non-linear functions of random variables, distribution transformations, and fundamental properties of functions of random variables. Section 2.3 revive the theory of stochastic processes and establishes the relations between a time- and spectral-representation. Finally Section 2.4 describes some basic information theoretic functions that are necessary for the development of the blind signal separation techniques.

2.1 Elements from Probability Theory

Probability theory and mathematical statistics form the basis of statistical signal processing, and the fields of probability theory and mathematical statistics are well founded in classical textbook as [36, 37, 30, 78]. To establish notation and terminology for the succeeding part of the thesis this section introduces elementary probability functions and the theory of moment and cumulant functions.

2.1.1 Probability Functions

The theory of probability deals with the description of random variables and the description is based on deterministic functions. In the following a number of such descriptive functions is introduced.

Distribution Functions The distribution function describes the probability for events of a random variable. A real valued function $F(x)$ is termed a distribution function if [65]

i F(x) is non-decreasing

ii F(x) is right continuous

iii $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

The distribution function has the interpretation: $P(x \leq x_0) = F(x_o) = \int_{-\infty}^{x_0} dF(x)$, i.e. the probability of an event of the random variable x is less than or equal to x_0. For continuous distribution functions the derivative of the distribution is termed the probability density function (pdf.) and is denoted $f(x)$. It follows from the characterization of the distribution function that $f(x) \geq 0 \; \forall \; x$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Integral Transformations Integral transformations of the distribution function generate new classes of functions that are useful in the theory of probability. A general presentation of integral transformations of the distribution function, can be written as

$$\int_{-\infty}^{\infty} K(t, x) dF(x)$$

where $K(t, x)$ is a suitable kernel, that is a function of the variable x and t [65]. In the following some important integral transformations are introduced.

First consider the kernel $K(x) = x^k$, provided the integral exist this kernel defines the transformation

$$m_k = \int_{-\infty}^{\infty} x^k dF(x) \tag{2.1}$$

and m_k is termed the algebraic moment of order k of $F(x)$. A number of related transformations is common; the kernel $K(x) = (x - \mu)^k$ where μ is the mean value, is termed the kth central moment; and the kernel $K(x) = |x|^k$ is termed the kth order absolute moment. The four first (central) moments are well-known: the first order moment is the mean value, the second order moment is the variance, the third order moment is the skewness, and the fourth order moment is the kurtosis. The existence of the moment functions is not assured for all distributions, and the existence should be established before the moment is used.

For the real variable t, the kernel $K(t, x) = \exp(tx)$ defines the transformation

$$\varphi(t) = \int_{-\infty}^{\infty} \exp(tx) dF(x) \tag{2.2}$$

Observe that

$$\left. \frac{d^k}{dt^k} \varphi(t) \right|_{t=0} = \int_{-\infty}^{\infty} x^k dF(x)$$

is the kth moment of $F(x)$, and for that reason $\varphi(t)$ is termed the moment generating function (mgf.) of $F(x)$. Clearly the existence of the moment generating function is not assured for all distributions, and the moment generating function exists if the integral in Equation 2.2 converges for t belonging to a defined interval, which contains the origin.

Finally consider the kernel $K(t, x) = \exp(jtx)$ for $t \in \mathbb{R}$ which defines the transformation

$$\phi(t) = \int_{-\infty}^{\infty} \exp(jtx) dF(x) \tag{2.3}$$

The function $\phi(t)$ is called the characteristic function (chf.) of $F(x)$. Every bounded function is integrable with respect to any distribution function over $(-\infty, \infty)$ and this assures the existence of the chf. defined in Equation 2.3 for every distribution function [65].

Characteristic Functions The characteristic function, defined in Equation 2.3, is the inverse Fourier transform of the pdf.. The Fourier transform pair $f(x) \leftrightarrow \phi(t)$ is unique and thus the properties for Fourier transform pairs apply to the pdf.-chf. pair. The chf. has the following properties [65]

i $\phi(0) = 1$

ii $|\phi(t)| \leq 1$

iii $\phi(-t) = \phi^*(t)$

Examples of $f(x) \leftrightarrow \phi(t)$ Pairs In Figure 2.1 four Fourier transform pairs $f(x) \leftrightarrow \phi(t)$ are depicted, and the corresponding pdf.s and chf.s are listed in Table 2.1.

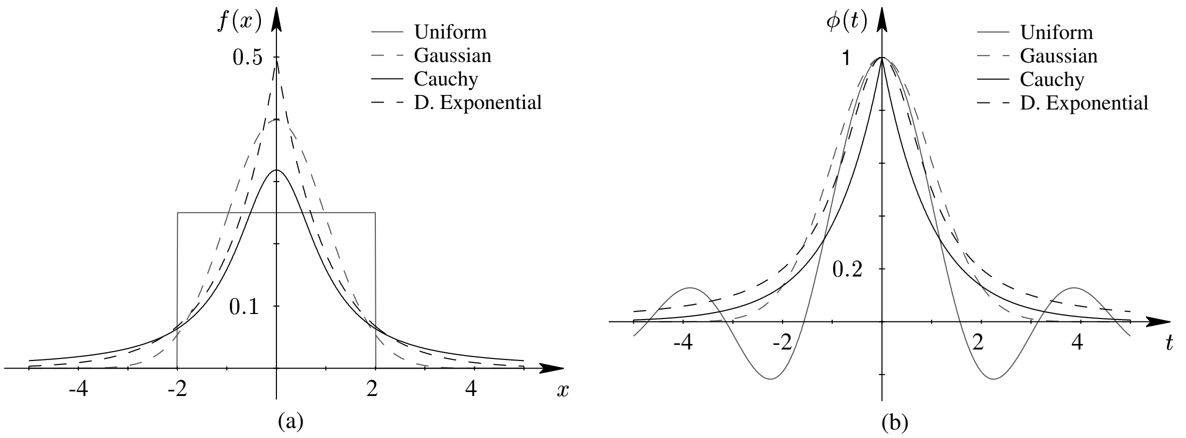


Figure 2.1 Depiction of four Fourier transform pairs. (a) Probability density functions. (b) Characteristic functions. The pdf.s and chf.s are parameterized in Table 2.1, and the plots in this figure are for the parameter values: $\mu = 0$, $r = 2$, $\sigma = 1$, $\gamma = 1$, and $\beta = 1$.

The pdf.s in Table 2.1 are all characterized by a location parameter μ and a scaling parameter. In general the chf. is a complex valued function but for the symmetric pdf. the corresponding chf. is a real valued function. The pdf.s in Table 2.1 are symmetric around the location parameter and from the chf. it appears that for $\mu \neq 0$ the chf. is a trivial version of the real valued chf. for $\mu = 0$.

2.1.2 Moments and Cumulants

For a distribution function $F(x)$ with finite integer Kth order moment the chf. $\phi(t)$ has the series expansion

$$\phi(t) = 1 + \sum_{k=1}^{K} \frac{m_k}{k!}(jt)^k + o(t^k) \tag{2.4}$$

where the integer kth moment is given in Equation 2.1. From the definition of the chf. in Equation 2.3 it appears that the kth moment is related to the kth derivative of the chf. as

$$(-j)^k \left. \frac{\partial^k}{\partial t^k} \phi(t) \right|_{t=0} = m_k \tag{2.5}$$

Even though the chf. is continuous the existence of moments is not assured.

Distribution	Probability density function	Characteristic function
Uniform	$f(x) = \begin{cases} \frac{1}{2r} & \text{if } \|x-\mu\| \le r \\ 0 & \text{if } \|x-\mu\| > r \end{cases}$	$\phi(t) = \frac{\sin(tr)}{tr} \exp(jt\mu)$
Gaussian	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\phi(t) = \exp\left(-\frac{\sigma^2 t^2}{2}\right) \exp(jt\mu)$
Cauchy	$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\mu)^2}$	$\phi(t) = \exp(-\gamma\|t\|) \exp(jt\mu)$
Double exponential	$f(x) = \frac{1}{2\beta} \exp\left(-\left\|\frac{x-\mu}{\beta}\right\|\right)$	$\phi(t) = \frac{1}{1+\beta^2 t^2} \exp(jt\mu)$

Table 2.1 Parameterization of the pdf.s and chf.s depicted in Figure 2.1.

The natural logarithm to the chf.

$$\psi(t) = \ln \phi(t) \tag{2.6}$$

is termed the cumulant generating function (cgf.). If the integer Kth moment of the distribution $F(x)$ exists, then the cgf. has the series expansion

$$\psi(t) = 1 + \sum_{k=1}^{K} \frac{c_k}{k!}(jt)^k + o(|t|^K) \tag{2.7}$$

where the coefficients c_k in the expansion are called cumulants. Thus the kth cumulant, c_k, is related to the kth derivative of the cgf. as

$$(-j)^k \left.\frac{\partial^k}{\partial t^k}\psi(t)\right|_{t=0} = c_k \tag{2.8}$$

Notice that while the cgf. of a pdf. is always defined the existence of the cumulants is not assured.

From the chf. and the cgf. the moments and cumulants are given by Equation 2.5 and 2.8. By the following example some of the characteristics of the distributions in Tabular 2.1 are exploited, and some more general characteristics are stated. For more formal derivations and proofs see [65].

Example In Table 2.2 the first four moments and cumulants of the distributions in Table 2.1 are listed. From the table a number of interesting observations appears which gives insight into the characteristics of the distributions and the properties of moments and cumulants in general.

For the Gaussian pdf. all the information about the distribution is contained in the first and second moments, and higher order moments provide no additional information. Furthermore cumulants of order higher than two are zero, cumulants of order greater than two can therefore be interpreted as a measure of the deviation from the Gaussian pdf.. The chf. of the Cauchy distribution is not differentiable in zero and therefore no moments or cumulants exist.

The first order moment and cumulant are the mean value. The second order cumulant is equal to the kth order central moment. Even order moments and cumulants describe the symmetric part of the pdf. whereas the odd order moments and cumulants describe the non-symmetric part. By comparing the coefficients in the series expansion of the chf. and the cgf. the relation between moments and cumulants can be obtained. The first four cumulants are related to the moments as

Distribution	Moments		Cumulants	
Uniform	$m_1 = \mu$	$, m_2 = \mu^2 + \frac{1}{3}r^2$	$c_1 = \mu$	$, c_2 = \frac{1}{3}r^2$
	$m_3 = \mu^3 + \mu r^2$	$, m_4 = \frac{1}{5}r^4 + \mu^4 + \mu^2 r^2$	$c_3 = 0$	$, c_4 = -\frac{2}{15}r^4$
Gaussian	$m_1 = \mu$	$, m_2 = \sigma^2 + \mu^2$	$c_1 = \mu$	$, c_2 = \sigma^2$
	$m_3 = 3\sigma^2\mu + \mu^3$	$, m_4 = 3\sigma^3 + 6\sigma^2\mu^2 + \mu^4$	$c_3 = 0$	$, c_4 = 0$
Cauchy	non-existent		non-existent	
Double exponential	$m_1 = \mu$	$, m_2 = 2\beta^2 + \mu^2$	$c_1 = \mu$	$, c_2 = 2\beta^2$
	$m_3 = 6\mu\beta^2 + \mu^3$	$, m_4 = 24\beta^4 + 12\mu^2\beta^2 + \mu^4$	$c_3 = 0$	$, c_4 = 12\beta^4$

Table 2.2 Moments and cumulants up to order 4 for the distributions listed in Table 2.1.

$$
\begin{array}{lcl}
c_1 = m_1 & , & c_2 = m_2 - m_1^2 \\
c_3 = m_3 - 3m_2 m_1 + 2m_1^3 & , & c_4 = m_4 - 4m_3 m_1 - 3m_2^2 + 12 m_2 m_1^2 - 6m_1^4
\end{array}
$$

The location parameter of a distribution in general has influence on all moments, whereas only the first order cumulant is dependent of the location.

Generalizing to Multivariate Distributions The theory is readily generalized to multivariate distributions. The chf. for the random vector $\mathbf{x}$ with distribution function $F(\mathbf{x})$ is

$$\phi(\mathbf{t}) = \int_{-\infty}^{\infty} \exp(j\mathbf{t}^T\mathbf{x}) dF(\mathbf{x}) \tag{2.9}$$

where $\mathbf{t} \in \mathbb{R}^N$; and the cgf. for the random vector is $\psi(\mathbf{t}) = \ln \phi(\mathbf{t})$. The series expansions also generalizes to the multidimensional chf. and cgf. resulting in the joint moments and cumulants

$$(-j)^k \left. \frac{\partial^k}{\partial t_1^{k_1} \ldots \partial t_N^{k_N}} \phi(\mathbf{t}) \right|_{\mathbf{t}=\mathbf{0}} = m_{\mathbf{k}} \qquad \text{and} \qquad (-j)^k \left. \frac{\partial^k}{\partial t_1^{k_1} \ldots \partial t_N^{k_N}} \psi(\mathbf{t}) \right|_{\mathbf{t}=\mathbf{0}} = c_{\mathbf{k}}$$

where $k = k_1 + \ldots k_N$ and $\mathbf{k} = [k_1, \ldots, k_N]$. For vectors it is sometimes convenient to operate with moments and cumulants on a functional form. For a random vector $\mathbf{x} = [x_1, \ldots, x_N]^T$ the following notation is introduced

$$m_{\mathbf{k}} = \text{mom}\left(x_1^{k_1}, \ldots, x_N^{k_N}\right) = \text{E}\left(x_1^{k_1} \ldots x_N^{k_N}\right) \qquad \text{and} \qquad c_{\mathbf{k}} = \text{cum}\left(x_1^{k_1}, \ldots, x_N^{k_N}\right)$$

In practice, at least in signal processing, moments and cumulants up to order 4 have practical applications.

2.1.3 Properties of Moments and Cumulants

The close relation between moments and cumulants is reflected in the fact that they share some important properties. As will appear in the following, the cumulants have some strong properties that are not valid for moments, and this is the reason why cumulants are generally preferred.

Moments and cumulants have the following scaling property [80, 72]

PROPERTY 2.1 *For random variables $x_1, \ldots, x_k$ and constants $a_1, \ldots, a_k$*

$$
\begin{aligned}
\text{mom}\,(a_1 x_1, \ldots, a_k x_k) &= a \text{ mom}\,(x_1, \ldots, x_k) \\
\text{cum}\,(a_1 x_1, \ldots, a_k x_k) &= a \text{ cum}\,(x_1, \ldots, x_k)
\end{aligned}
$$

where $a = \prod_{l=1}^{k} a_l$.

Moments and cumulants have the following symmetry property [80, 72]

PROPERTY 2.2 *For random variables* $x_1, \dots, x_k$

$$\begin{aligned} \text{mom}\,(x_1, \dots, x_k) &= \text{mom}\,(x_{l_1}, \dots, x_{l_k}) \\ \text{cum}\,(x_1, \dots, x_k) &= \text{cum}\,(x_{l_1}, \dots, x_{l_k}) \end{aligned}$$

where $(l_1, \dots, l_k)$ *is a permutation of* $(1, \dots, k)$.

This means that one can interchange the arguments of the moment/cumulant without changing the value of the moment/cumulant.

Sums of Random Variables A well known result from probability theory is that the pdf. of sums of independent variables is the convolution of the pdf.s of the variables in the sum [36]. Consider the sum

$$x = \sum_{n=1}^{N} x_n$$

where $x_1, \dots, x_N$ are independent random variables. Let $f_1(x), \dots, f_N(x)$ denote the pdf.s of the random variables, then the pdf. of the sum can be written on the compact form

$$f_x(x) = f_1(x) * \cdots * f_N(x)$$

where $*$ denotes convolution, $f_1(x) * f_2(x) = \int_{-\infty}^{\infty} f_1(z) f_2(x - z) dz$. Let $\phi_1(t), \dots, \phi_N(t)$ denote the chf.s of the random variables, then the chf. of the sum is

$$\phi_x(t) = \prod_{n=1}^{N} \phi_n(t)$$

This follows from the definition of the chf. and the convolution property for the Fourier transform. Furthermore, let $\psi_1(t), \dots, \psi_N(t)$ denote the cgf.s of the random variables, then the cgf. of the sum is

$$\psi_x(t) = \sum_{n=1}^{N} \psi_n(t)$$

From these equations the following property for cumulants appears.

PROPERTY 2.3 *Cumulants of a sum of statistically independent quantities equal the sum of the cumulants of the individual quantities. If the random variables* $x_1, \dots, x_N$ *are independent of the random variables* $y_1, \dots, y_N$ *then*

$$\text{cum}\,(x_1 + y_1, \dots, x_N + y_N) = \text{cum}\,(x_1, \dots, x_N) + \text{cum}\,(y_1, \dots, y_N)$$

This strong property has given name to the cumulants. Notice that Property 2.3 does not apply to moments in general.

If the random variables $x_1, \dots, x_k$ can be divided into any two groups, $x_1, \dots, x_n$ and $x_{n+1}, \dots, x_k$, which are statistically independent, then the chf. can be written as

$$\phi(t_1, \dots, t_k) = \phi(t_1, \dots, t_m)\ \phi(t_{m+1}, \dots, t_k)$$

and the cgf. as

$$\psi(t_1,\ldots,t_k) = \psi(t_1,\ldots,t_m) + \psi(t_{m+1},\ldots,t_k)$$

From this it appears that the kth-order cumulant cum $(x_1,\ldots,x_k) = 0$ whereas for the kth-order moment in general mom $(x_1,\ldots,x_k) \neq 0$.

2.2 Functions of Random Variables

Non-linear functions of random variables are important to blind techniques in signal processing. This section reviews some fundamental properties of functions of random variables and these properties will be used in Chapter 4. Furthermore a simple empirical method for density transformations is proposed.

2.2.1 Functions of One Random Variable

Let x be a random variable with pdf. $f_x(x)$ and consider a function $y = g(x)$. If x is a random variable and g is a non-constant function then y is also a random variable. Figure 2.2 is a graphical depiction of the pdf. transformation of $f_x(x)$ with $g(x)$ to the pdf. of the transformed variable $f_y(y)$.

From the illustration in Figure 2.2 it appears that in the interval $[y_1 < y < y_4]$ the pdf. $f_y(y)$ is a sum over three partitions of $f_x(x)$, outside the interval it is a sum over one partition; e.g. the probability mass

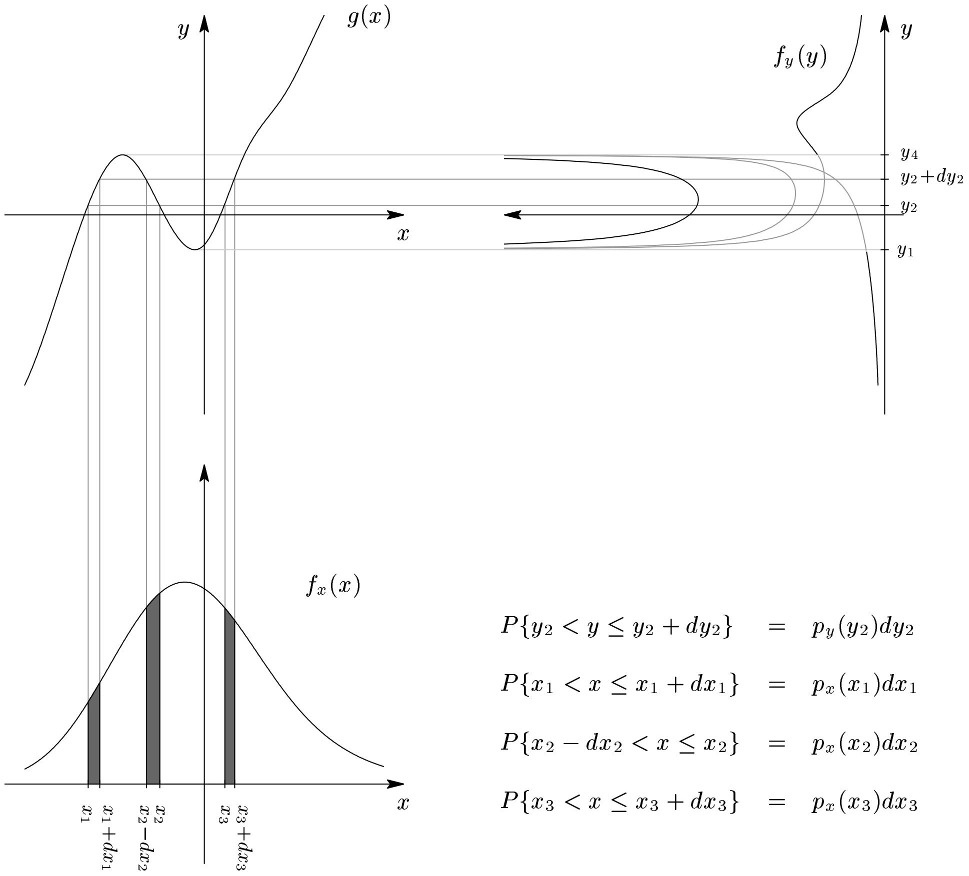


Figure 2.2 Depiction of relation between the pdf.s $f_x(x)$ and $f_y(y)$ for the function $y = g(x)$.

of the random variable y in the interval $[y_2, y_2 + dy_2]$ is the sum of the probability mass in the intervals $[x_1, x_1 + dx_1]$, $[x_2, x_2 + dx_2]$, and $[x_3, x_3 + dx_3]$ from $f_x(x)$. In a more formal form we have

$$f_y(y_2) = f_x(x_1)\frac{dx_1}{dy_2} + f_x(x_2)\frac{dx_2}{dy_2} + f_x(x_3)\frac{dx_3}{dy_2}$$

From the graphical construction of the pdf. in Figure 2.2 a general form of the transformation from $f_x(x)$ to $f_y(y)$ is obtained

$$f_y(y) = f_x(x_1)\left|\frac{dx}{dg(x_1)}\right| + \cdots + f_x(x_n)\left|\frac{dx}{dg(x_n)}\right| + \ldots \tag{2.10}$$

where x_n is the roots in $y = g_x(x)$. For a more formal derivation of this result see [78].

2.2.2 Transformations of Probability Density Functions

The density transformation problem is to determine a function $y = g(x)$, given the probability density $f_x(x)$, that transforms the random variable x into the variable y which follows the desired pdf. $f_y(y)$. To avoid discontinuities in $f_y(y)$ the function $g(x)$ is required to be differentiable and with a non-zero derivative in the domain of support of $f_x(x)$. From Equation 2.10 the pdf. of the random variable y is

$$f_y(y) = f_x(x)\left|\frac{dx}{dg(x)}\right| \tag{2.11}$$

In Figure 2.3(a) a density transform is depicted.

Transformation of Uniform Probability Density Function Let x be a random variable with a uniform pdf.. Consider the inverse density transformation $x = g^{-1}(y)$ that transforms y into x, where y follows an arbitrary pdf. $p_y(y)$. Using Equation 2.11, the solution to this is given by the differential equation $p_y(y) = \frac{dx}{dy}$ which has the solution

$$x = \int f_y(y)dy$$

The transformation that transforms a uniform distributed variable x into a variable y, that follows the distribution $f_y(y)$, is therefore the inverse function to the distribution of y. The transformation is depicted in Figure 2.3(b).

Thus, transformation between an arbitrary continuous probability distribution to or from a uniform probability distribution, is given by the probability density function or the inverse probability density function respectively.

Empirical Transformation to Uniform Distribution Transformation between pdf.s in general requires that the source distribution is known a priori or that the source distribution is estimated. Here we propose a different approach which is numerically very robust and is computational inexpensive.

Let x be a random variable, and let $x_1, x_2, \ldots, x_N$ be N observations of the variable. An empirical transformation from an unknown arbitrary pdf. to the uniform pdf. is readily obtained by sorting the observations. Sort the N observations, by the index m, such that

$$x_{\cdot,1} \leq x_{\cdot,2} \leq \cdots \leq x_{n,m} \leq \cdots \leq x_{\cdot,N}$$

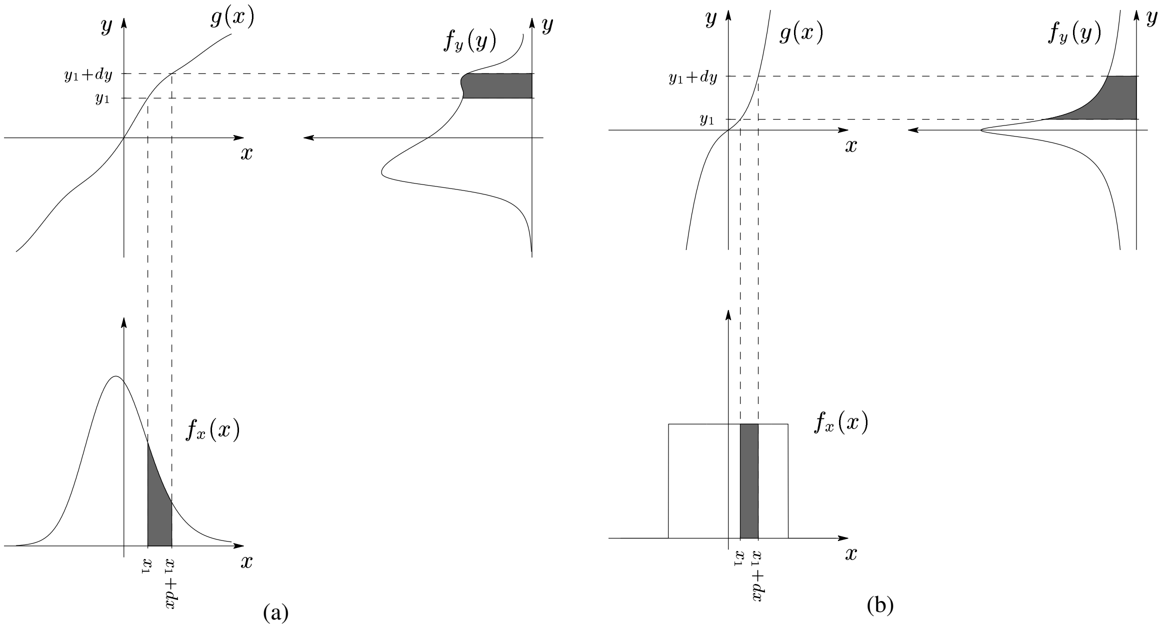


Figure 2.3 Depiction of density transformations. (a) Transformation between arbitrary continuous pdf.s. (b) Transformation of uniform pdf..

and assign y_n the value

$$y_n = \frac{2m - Nr}{N} + \mu$$

where m is the index that corresponds to the nth observation of x. Now, the random variable y follows a uniform pdf., where the variables r and μ are defined in Table 2.1. Notice that the transformation is non-parametric. The empirical transformation is only dependent on the observations, and the hidden function $g(x)$ can be interpreted as the inverse to the instantaneous estimate of the cumulated histogram of x.

The transformation is readily modified to transform to an arbitrary pdf. by applying the inverse of the desired distribution function to the uniform transformed variable. A novel practical application for the empirical distribution transformation is shown in [54] included in Appendix A.1, where the transformation is used in the adaptation of an adaptive filter.

2.2.3 Functions of Multivariate Random Variables

The result for univariate random variables in Equation 2.10 generalizes to the multivariate random variables in the following way. Consider a function $\mathbf{y} = g(\mathbf{x})$ where $\mathbf{x} = [x_1, \ldots, x_N]^T$ and $\mathbf{y} = [y_1, \ldots, y_N]^T$ are random vectors and g is a general function of the form $y_1 = g_1(\mathbf{x})$, … ,$y_N = g_N(\mathbf{x})$. The joint pdf. of $\mathbf{y}$ is

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{f_{\mathbf{x}}(\mathbf{x}_1)}{|\det J(\mathbf{x}_1)|} + \cdots + \frac{f_{\mathbf{x}}(\mathbf{x}_n)}{|\det J(\mathbf{x}_n)|} + \cdots + \tag{2.12}$$

where $\mathbf{x}_n$ is the nth real root of the transformation g and $J(\mathbf{x})$ is the Jacobian of g, for a proof see [78].

Linear Transformation Consider the linear transformation $\mathbf{y} = \mathbf{A}\mathbf{x}$. From Equation 2.12 the joint pdf. of $\mathbf{y}$ is

$$f_{\mathbf{y}}(\mathbf{y}) = |\det \mathbf{A}|^{-1} f_{\mathbf{x}}(\mathbf{x}) \tag{2.13}$$

In the scope of this work it is natural to think of the linear transformation as an instantaneous mixture of signals, and Equation 2.13 relates the pdf.s of the signals and the mixture of the signals.

Observe that if $\mathbf{x}$ is a Gaussian vector with pdf.

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{2.14}$$

then any linear transformation is also Gaussian with mean vector $\mathbf{A}\boldsymbol{\mu}$ and covariance matrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$. From this observation it appears that for a mixture of Gaussian signals, the mixture matrix, $\mathbf{A}$, can not be identified based on the covariance matrix.

2.3 Stochastic Processes and Linear Filters

One of the most important models in statistical signal processing is a weakly stationary stochastic process[1], a process whose first and second moments only depend on the time difference in which observations are made. The importance of the weak stationarity is due to the fact that it implies that a Fourier (harmonic) analysis of both the process and the covariance of the process can be carried out. Therefore both a time domain and a frequency domain representation of the process are valid. This section establishes the fundamentals from the theory of discrete-time stationary stochastic processes, and some implications for the blind signal separation problem are considered.

2.3.1 Discrete-Time Stationary Stochastic Processes

Let $x(n)$ represent a real valued stochastic process in time, where n is the discrete time variable. The stochastic process is weak stationary if the second order moment function

$$m_2(l) = \mathrm{E}\left(x(n)x(n+l)\right) \quad , \quad \text{for all } n$$

only depends on the time difference l. For stationary stochastic processes the (central) second order moment function is often referred to as the autocorrelation function, $r_x(l) = m_2(l) - m_1^2$. The stochastic process is (strictly) stationary if the joint distribution function of any subset of observations do not change over time, e.g.

$$F(x(n), x(n+l)) = F(x(m), x(m+l))$$

for all time indexes n and m and all time differences l.

If $x(n)$ is a real stationary stochastic process and the moment up to order k exists, then the moment and cumulant functions will only depend on the time differences $l_1, \ldots, l_{k-1}$. The moment and cumulant functions of such a stochastic process are

$$\begin{aligned} m_{k,x}(l_1, \ldots, l_{k-1}) &= \mathrm{E}\left(x(n)x(n+l_1)\ldots x(n+l_{k-1})\right) \\ c_{k,x}(l_1, \ldots, l_{k-1}) &= \mathrm{cum}\left(x(n), x(n+l_1), \ldots, x(n+l_{k-1})\right) \end{aligned}$$

respectively. The general relation between moments and cumulants are derived in [63].

[1] *Weakly* stationary and *wide-sense* stationary are used interchangeably in the literature.

Notice that the condition for (strict) stationarity imply that all moments are invariant in time, whereas the conditions for weak stationarity only rely on that the first and second order moment is invariant in time. Since the existence of the second order moment function are not assured for stochastic processes in general, stationarity does not necessarily imply weakly stationarity and vice versa. Often weakly stationarity is considered as a weaker condition than stationarity, this is not true in general and in Section 3.2.3 classes of stationary processes that are not weakly stationary are considered.

ARMA-processes and Linear Filters A causal linear filter is characterized by the impulse response $h(n)$ and has a corresponding one-sided z-transform $H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}$. If the poles of $H(z)$ are inside the unit circle the transfer function is stable. If all zeros of $H(z)$ are inside the unit circle the linear filter is termed minimum phase, if all zeros are outside the unit circle the linear filter is termed maximum phase, and linear filters with zeros inside and outside the unit circle are termed mixed phase [77].

One of the most important stochastic processes in signal processing is the ARMA(K, L)-process

$$x(n) + \sum_{k=1}^{K} a_k x(n-k) = \sum_{l=0}^{L} b_l u(n-l)$$

where $u(n)$ is an iid. stochastic process and is termed the excitation process, and the coefficients a_k and b_l are parameters of the process. For $L = 0$ the process is termed an AR(K)-process and for $K = 0$ the process is termed an MA(K)-process. If the excitation process is iid. Gaussian then the process is termed a Gaussian process, and consequently any finite subset of observations is jointly Gaussian. The ARMA-process can also be characterized by the z-transformed $H(z) = B(z)/A(z)$ where $A(z) = 1 + \sum_{k=1}^{K} a_k z^{-k}$ and $B(z) = \sum_{l=0}^{L} a_l z^{-n}$.

The Cumulant Function of a Stochastic Process The kth-order cumulant function of a stationary stochastic process $x(n)$ is related to the impulse response of the process as[2]

$$c_{k,x}(l_1, l_2, \ldots, l_{k-1}) = c_{k,u} \sum_{n=0}^{\infty} \prod_{m=0}^{k-1} h(n + l_m) \tag{2.15}$$

where $c_{k,u}$ is the kth-order cumulant of the excitation process $u(n)$.

Using Equation 2.15 the third order cumulant function of the process is

$$c_{3,x}(l_1, l_2) = c_{3,u} \sum_{n=0}^{\infty} h(n)h(n + l_1)h(n + l_2)$$

where $c_{3,u}$ is the third order cumulant function of the excitation process. Assuming that $h(1) = 1$ and combining the two cumulant points $c_{3,x}(l_1, l_2) = c_{3,u}h(l_1)h(l_2)$ and $c_{3,x}(l_1, 0) = c_{3,u}h(l_1)$ the following closed form of the kth coefficient in the MA-process is obtained [39]

$$h(k) = \frac{c_3^y(l_1, k)}{c_3^y(l_1, 0)} \quad , \quad \text{for} \quad k = 1, \ldots, K \tag{2.16}$$

Even though Equation 2.16 is not the best estimator of the general MA impulse response it shows that the cumulant reveals, within a scale factor, the impulse response. This is further demonstrated in the following example.

[2]This equation is often referred to as the Barlett-Brillinger-Rosenblatt summation formula, for references see [39].

Example: Minimum, Maximum and Mixed Phase Systems For an MA(2)-process with real zeros in the transfer function, there are 3 different realizations[3]. The realization with both zeros inside the unit circle is referred to as the minimum phase realization, with both zeros outside the unit circle the realization is referred to as a maximum phase realization, and for one zero inside and one zero outside the unit circle the realization are referred to as a mixed phase realization. In Table 2.3 the transfer function and impulse response for the minimum, maximum and two different mixed phase realizations are given, and the corresponding zero location in the z-plane is depicted.

The autocorrelation of the four different transfer functions is identical, and consequently the autocorrelation function can not distinguish between the different realizations. Assuming that the excitation of the MA(2)-process is non-Gaussian with the third cumulant different from zero, the third order cumulants of the process are also different from zero. Calculating the third order cumulants and combining the cumulants as in Equation 2.16 reveals the filter coefficients of the transfer function. This is shown in the bottom part of Table 2.3, and this demonstrates that the cumulants provide information that is not available from the second order moments.

Covariance and Non-Linear Functions The covariance or correlation function is usually used to describe the temporal structure of signals. As illuminated in the previous example, second order moment functions are in general insufficient in blind estimation and thus it is necessary to use higher order statistics or non-linear estimation functions. Often higher order moments and non-linear estimation functions can be considered as a second order moment function of a non-linear function of the signal, and therefore it is instructive to consider the so-called *Bussgang* property.

For jointly Gaussian variables x and y the joint probability density function depends only on the covariance function, and has the explicit expression

$$f_{xy}(x,y) = \frac{1}{2\pi\sqrt{1-r_{xy}}} \exp\left(-\frac{x^2+y^2-2r_{xy}xy}{2(1-r_{xy})}\right)$$

Let $v = g(y)$ where g is a non-linear function of the random variable y. The covariance function of x and a non-linear function of y is [16]

$$r_{xv} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\ g(y)\ f_{xy}(x,y)dx\ dy = \frac{r_{xy}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y\ g(y)\exp\left(-\frac{y^2}{2}\right)\ dy = Kr_{xy} \tag{2.17}$$

where K is a constant depending on the function g. Thus, the correlation between two Gaussian variables is proportional to the correlation function of one of the Gaussian variable and a non-linear transformation of the other, this property is known as the Bussgang property. Clearly, the function g must not be an even function, and only the odd part contributes to the constant K.

2.3.2 Spectral Representation

Consider a discrete-time weakly stationary zero-mean stochastic process $x(n)$ with covariance function $r_x(l)$. The covariance function has the Fourier representation

$$r_x(l) = \int_{-\pi}^{\pi} e^{j\omega l} dG_x(\omega) \tag{2.18}$$

with $G_x(-\pi) = 0$ and $G_x(\pi) = r_x(0)$. This representation is termed the *spectral representation*, and $G_x(\omega)$ is termed the spectral distribution function.

[3]This example is inspired by [72, §2.4.3].

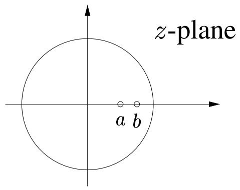
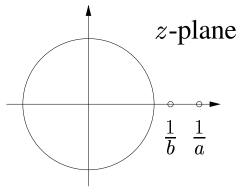
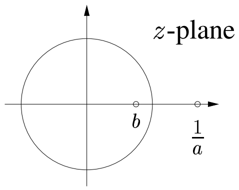
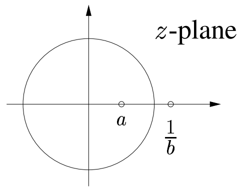
	Minimum Phase	Maximum Phase	Mixed Phase I	Mixed Phase II
Transfer function				
$H(z)$	$(1-az^{-1})(1-bz^{-1})$	$(1-az)(1-bz)$	$(1-az)(1-bz^{-1})$	$(1-az^{-1})(1-bz)$
Causal impulse response				
$h(0)$	1	ab	a	b
$h(1)$	$-(a+b)$	$-(a+b)$	$1+ab$	$1+ab$
$h(2)$	ab	1	$-b$	$-a$
Zero location				
Autocorrelation				
$c_2(0)$	$1+a^2b^2+(a+b)^2$			
$c_2(1)$	$-(a+b)(1+ab)$			
$c_2(2)$	ab			
Third-order cumulants				
$c_3(0,0)$	$1-(a+b)^3+a^3b^3$	$1-(a+b)^3+a^3b^3$	$(1+ab)^3-a^3-b^3$	$(1+ab)^3-a^3-b^3$
$c_3(1,1)$	$(a+b)^2-(a+b)a^2b^2$	$-(a+b)^2+ab(a+b)^2$	$-a(1+ab)^2+(1+ab)^2b$	$-b(1+ab)^2+(1+ab)a^2b$
$c_3(2,2)$	a^2b^2	ab	ab^2	ba^2
$c_3(1,0)$	$-(a+b)+ab(a+b)^2$	$(a+b)^2-(a^2+b)^2$	$a^2(1+ab)-(1+ab)^2b$	$b^2(1+ab)-(1+ab)^2a$
$c_3(2,0)$	ab	a^2b^2	$-a^2b$	$-b^2a$
$c_3(2,1)$	$-(a+b)ab$	$-(a+b)ab$	$ab(1+ab)$	$ab(1+ab)$
Combination of cumulants				
$h(1)=\dfrac{c_3(2,1)}{c_3(2,0)}$	$-(a+b)$	$\dfrac{-(a+b)}{ab}$	$\dfrac{1+ab}{a}$	$\dfrac{1+ab}{b}$
$h(2)=\dfrac{c_3(2,2)}{c_3(2,0)}$	ab	$\dfrac{1}{ab}$	$-\dfrac{b}{a}$	$-\dfrac{b}{a}$

Table 2.3 MA(2)-process with parameters $a, b \in \mathbb{R}$, $0 < a < 1$, $0 < b < 1$. Example of a closed form identification of a minimum, maximum and mixed phase systems impulse responses using Equation 2.16.

If G_x is differentiable the derivative $dG_x(\omega)/d\omega$ is termed the spectral density function [80]. The spectral density function is a density function in the probabilistic sense (a pdf.) and it can be interpreted as the average power of $x(n)$ in the frequency interval $[\omega\ ;\ \omega + d\omega]$. The discrete-time weakly stationary stochastic process has now the Fourier transform pairs

$$r_x(l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega l} S_x(\omega) d\omega \qquad \longleftrightarrow \qquad S_x(\omega) = \sum_{l=-\infty}^{\infty} r_x(l) e^{-j\omega l} \tag{2.19}$$

and $S_x(\omega)$ is termed the power spectrum of the stochastic process[4].

For Gaussian processes, the first and second order moment functions is a complete description of the probability structure, and consequently the time domain and the frequency representations in Equation 2.19 are sufficient and equivalent.

A linear filter characterized by the discrete transfer function $H(z)$ has the spectral representation $H(\omega) = H(z)|_{z=e^{-j\omega}}$ and is termed the *frequency response* of the filter. The spectral density function of the output of the filter is

$$S_x(\omega) = |H(\omega)|^2 S_u(\omega) \tag{2.20}$$

where $S_u(\omega)$ is the spectral density function of the input signal. It appears, analogous to the results in Section 2.3.1, that the second order moment function does not contain information about the phase information in the linear filter.

Cumulant Spectra As for the covariance function, the cumulant functions have a frequency domain counterpart termed *cumulant spectra*. The kth order cumulant spectrum is defined as the $k-1$ dimensional Fourier transform of the cumulant function

$$C_{k,x}(\omega_1, \ldots, \omega_{k-1}) = \sum_{l_1=-\infty}^{\infty} \cdots \sum_{l_{k-1}=-\infty}^{\infty} c_{k,x}(l_1, l_2, \ldots, l_{k-1}) \exp(-j(\omega_1 l_1 + \cdots + \omega_{k-1} l_{k-1})$$

[71]. For a linear filter with frequency response $H(\omega)$ the kth order cumulant spectrum of the output signal is

$$C_{k,x}(\omega_1, \ldots, \omega_{k-1}) = H(\omega_1) \ldots H(\omega_{k-1}) H^*(\omega_1 + \cdots + \omega_{k-1}) C_{k,u}(\omega_1, \ldots, \omega_{k-1}) \tag{2.21}$$

where $C_{k,u}(\omega_1, \ldots, \omega_{k-1})$ is the cumulant spectrum of the input signal [72]. This is a generalization of the power spectral representation in Equation 2.20. It appears that the cumulant spectrum of order $k > 2$, in contrary to the power spectrum, carries the phase information of $H(\omega)$. Notice that Equation 2.21 is the spectral counterpart of Equation 2.15.

2.4 Information Theory

In Section 1.2.2 an intuitive introduction to the blind signal separation problem was given using an abstract concept of measuring information. In this section the concept of measuring information is further elaborated and some other basic information theoretic functions, that are necessary for the development of the blind signal separation techniques described in Section 4.1, are introduced.

[4]The stationary stochastic process itself has also a Fourier representation that corresponds to the spectral representation of the covariance function. The Fourier representation can be written as $x(n) = \int_{-\pi}^{\pi} e^{j\omega n} dZ(\omega)$, where $Z(\omega)$ is a process of orthogonal increments, see [80].

2.4.1 Measures of Information

Entropy For a random variable or vector with a continuous pdf. it is useful to introduce an information theoretic measure termed the *entropy*. The entropy is a real-valued function of the pdf. that characterizes the uncertainty or randomness of the corresponding random variable.

The entropy of a random vector $\mathbf{x}$ with continuous pdf. $f_{\mathbf{x}}(\mathbf{x})$ is defined as[5]

$$H(\mathbf{x}) = -\int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}) \ln f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \tag{2.22}$$

The integration extends only over the region where $f(\mathbf{x}) \neq 0$ because $f(\mathbf{x}) \ln f(\mathbf{x}) = 0$ for $f(\mathbf{x}) = 0$, [78]. The entropy for continuous random variables is a relative measure and can take any values from $-\infty$ to ∞. Remark that the existence of the integral in Equation 2.22 is not assured and the entropy is thus not defined for all distributions. For bounded distributions the uniform distribution has the largest entropy, and for unbounded distributions with equal second order moments the Gaussian distribution has the largest entropy [29].

For two random variables x and y with joint pdf. $f(x, y)$, the *conditional entropy* is

$$H(x|y) = -\int_{-\infty}^{\infty} f(x, y) \ln f(x|y) dx dy \tag{2.23}$$

Using Bayes rule, $f(x|y) = f(x, y)/f(y)$, the conditional entropy can be written as

$$H(x|y) = H(x, y) - H(y)$$

Kullback-Leibler Distance / Relative Entropy The relative entropy or Kullback-Leibler distance is a measure of distance between two pdf.s. The Kullback-Leibler distance between the pdf.s f_1 and f_2 is denoted $D(f_1 \| f_2)$ and is defined as

$$D(f_1 \| f_2) = \int_{-\infty}^{\infty} f_1(\mathbf{x}) \ln \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} d\mathbf{x} \tag{2.24}$$

[29]. The Kullback-Leibler distance is a positive semidefinite function which is zero only if f_1 is equal f_2.

The Kullback-Leibler distance is not a true distance measure between distributions since it is not symmetric and does not satisfy the triangle inequality. Nonetheless, it is useful to think of the Kullback-Leibler distance as a *distance* between distributions.

Mutual Information The mutual information is a measure of the dependence of the elements in a random vector and is defined as the Kullback-Leibler distance between the joint pdf. and the product of the marginal pdf.s. For a random vector $\mathbf{x} = [x_1 \ \dots \ x_N]^T$ with joint pdf. $f_{\mathbf{x}}(\mathbf{x})$ and marginal pdf.s. $f_n(x_n)$ for $n = 1, \dots, N$, the mutual information is

$$I(\mathbf{x}) = D\left(f_{\mathbf{x}}(\mathbf{x}) \middle\| \prod_{n=1}^{N} f_n(x_n) \right) = \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}) \ln \frac{f_{\mathbf{x}}(\mathbf{x})}{\prod_{n=1}^{N} f_n(x_n)} d\mathbf{x} \tag{2.25}$$

[5] Entropy is defined in a different way for discrete pdf.s. In the literature, to distinguish between entropy for discrete and continuous distributions, the definition in Equation 2.22 is sometimes referred to as the *differential entropy*.

If the elements in $\mathbf{x}$ are independent then the joint pdf. is equal to the product of the marginal pdf.s, and from Equation 2.25 it appears that the mutual information $I(\mathbf{x}) = 0$. Mutual information is a positive semidefinite function and is zero if and only if the elements are independent.

From the definition of mutual information in Equation 2.25 and the definition of the entropy in Equation 2.22 the relation

$$H(\mathbf{x}) = \sum_{n=1}^{N} H(x_n) - I(\mathbf{x})$$

between the entropy and the mutual information appears [29].

Negentropy Negentropy is defined as the Kullback-Leibler distance between the the pdf. $f(\mathbf{x})$ and the Gaussian pdf. $f_G(\mathbf{x})$ with same mean and variance as $f(\mathbf{x})$ [29]

$$D\left(f \,\|f_G\right) = \int_{-\infty}^{\infty} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{f_G(\mathbf{x})} d\mathbf{x} \tag{2.26}$$

2.4.2 Functions of Random Variables

The generative model for mixtures of signals is a linear transformation and many of the blind separation methods are based on non-linear functions. Therefore a fundamental insight into the information theoretic perspectives of functions of random variables is necessary for the development of blind signal separation methods.

First recall that if the random variables u and x are independent, then for any non-linear function g the random variable $y = g(u)$ is also independent of x [78].

Transformations of a Random Vector Consider the random vector $\mathbf{y} = g(\mathbf{x})$ where g is an invertible function of the random vector $\mathbf{x}$. From Equation 2.12 the pdf. of $\mathbf{y}$ is $f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{x}}(\mathbf{x})/|\det J(\mathbf{x})|$ where $J(\mathbf{x})$ is the Jacobian of g and $f_{\mathbf{x}}(\mathbf{x})$ is the pdf. of $\mathbf{x}$. Using Equation 2.22 the entropy of $\mathbf{y}$ is [78]

$$H(\mathbf{y}) = H(\mathbf{x}) + \int_{-\infty}^{\infty} f(\mathbf{x}) \ln |\det J(\mathbf{x})| d\mathbf{x} \tag{2.27}$$

Linear Transformations For the linear transformation $\mathbf{y} = \mathbf{A}\mathbf{x}$ the entropy is readily obtained from Equation 2.27

$$H(\mathbf{y}) = H(\mathbf{x}) + \ln |\det \mathbf{A}|$$

It follows that the entropy is invariant to orthogonal transformations, since the determinant of orthogonal transformations equal 1.

Example: Kullback-Leibler Distance and Linear Transformations It is interesting, having the blind signal separation problem in mind, to consider the Kullback-Leibler distance measure between a random vector and a linear transformation of the vector. Let $\mathbf{x} \in \mathbb{R}^2$ be a random vector with independent double exponential distributed elements, $\mathbf{A} \in \mathbb{R}^{2\times 2}$ a linear transformation matrix, and $\mathbf{y} = \mathbf{A}\mathbf{x}$. From Equation 2.13 it is known that the pdf. of $\mathbf{y}$ is $f_{\mathbf{y}}(\mathbf{y}) = |\det \mathbf{A}|^{-1} f_{\mathbf{x}}(\mathbf{x})$. In this example it is usefull to parameterize the linear transformation matrix as $\mathbf{A} = [\cos(\theta_1) \quad \sin(\theta_2)\ ;\ \sin(\theta_1) \quad \cos(\theta_2)]$, where θ_1 and θ_2 are referred to as the directions of the columns in $\mathbf{A}$.

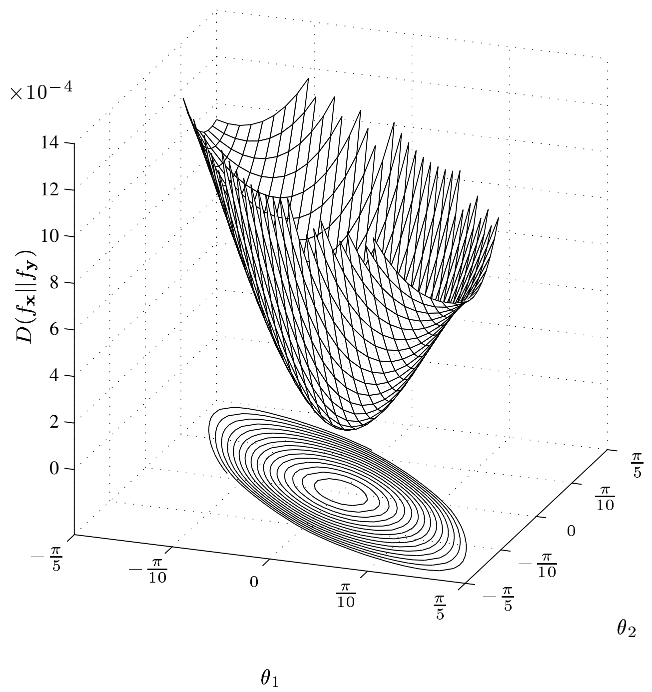


Figure 2.4 Kullback-Leibler distance versus the direction of the columns in a 2×2 mixture matrix.

In Figure 2.4 the Kullback-Leibler distance is depicted as a function of the directions of the columns in $\mathbf{A}$. For $\theta_1 = \theta_2 = 0$ the linear transformation matrix, $\mathbf{A}$, is the identity matrix and thus $D\left(f_{\mathbf{x}}\|f_{\mathbf{y}}\right) = 0$, and for $\theta_1 \neq \theta_2$ the Kullback-Leibler distance $D\left(f_{\mathbf{x}}\|f_{\mathbf{y}}\right) > 0$. These observations are consistent with the interpretation of $D\left(f_{\mathbf{x}}\|f_{\mathbf{y}}\right)$ as a distance measure. For $\theta_1 = -\theta_2$ the linear transformation matrix is orthogonal and from the figure it appears the Kullback-Leibler distance measure is very sensitive to non-orthogonality in $\mathbf{y}$. The "valley" of $D\left(f_{\mathbf{x}}\|f_{\mathbf{y}}\right)$ is the Kullback-Leibler distance on the manifold of orthogonal mixture matrices.

In the scope of this thesis it is instructive to consider $f_{\mathbf{x}}(\mathbf{x})$ as the pdf. model of the independent source signals, $\mathbf{A}$ as an instantaneous mixture matrix, and $f_{\mathbf{y}}(\mathbf{y})$ as the generative model.

Example: Entropy of a Non-linear function of a Gaussian Variable To investigate the entropy of a bounded random variable it is useful to consider a bounded non-linear function of a Gaussian variable. Let $x \sim \mathrm{N}(0,1)$ and $y = \tanh(ax)$ where a is a scaling parameter. The pdf. of the random variable x is depicted in Figure 2.5(c). In Figure 2.5(b) the pdf. of the random variable y is depicted for three different values of a, and the corresponding non-linear function is depicted in Figure 2.5(a). The entropy $H(y)$ as a function of the scaling parameter a is depicted in Figure 2.5(d). It is interesting to observe that $H(y)$ has maximum where $f_y(y)$ is closest to a uniform pdf., and this observation is consistent with the characterics of the entropy as discussed in Section 2.4.1.

2.5 Summary

In this chapter the basis of statistical signal processing from probability theory, mathematical statistics and information theory are reviewed. It is demonstrated that second order statistics in general are insufficient for blind estimation methods. To overcome the limitations of second order statistics the theoretical

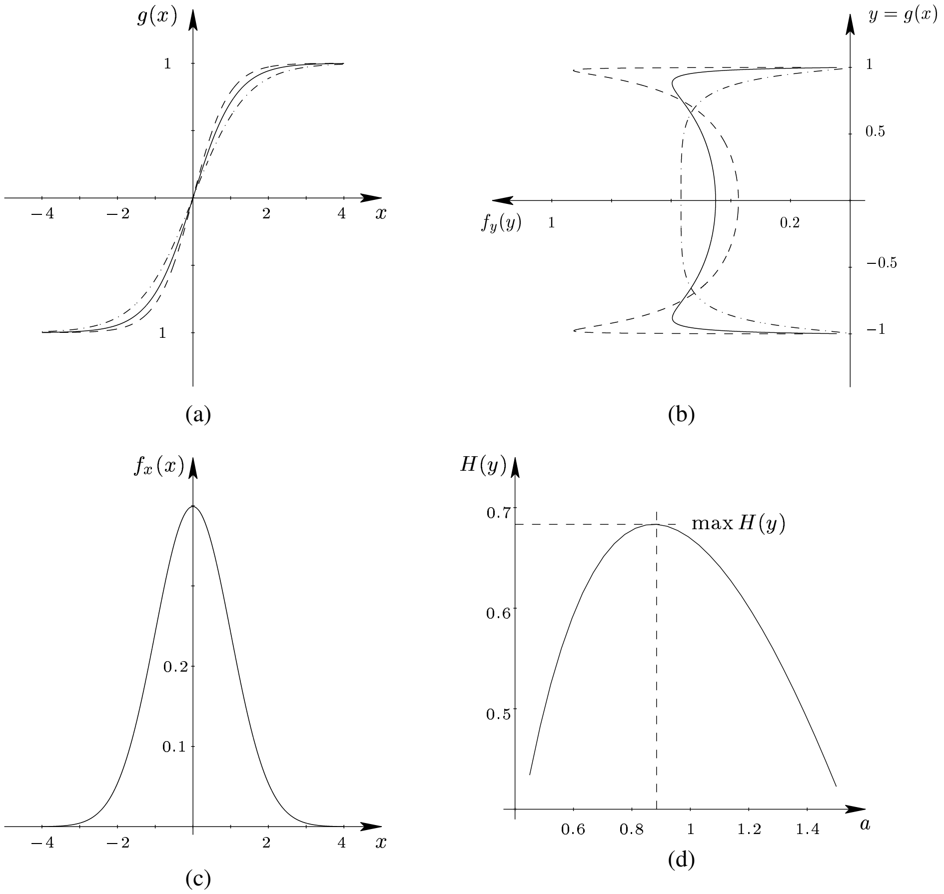


Figure 2.5 (a) Non-linear function, $g(x) = \tanh(ax)$ for $a = 0.7034$ (dash-dot), $a = 0.8845$ (solid) and $a = 1.0655$ (dashed). (b) Pdf. of transformed variable, $y = g(x)$ (c) Gaussian pdf., $\sigma^2 = 1, \mu = 0$ (d) Entropy of y versus the scaling parameter.

foundation for moments, cumulants and non-linear estimation functions are established. Thereby, the fundamentals for operating with heavy tail distributed signals and blind signal separation is established.

A graphical interpretation of non-linear functions of random variables are presented; and a novel empirical transformation method transforming a signal so that it follows an arbitrary distribution is proposed.

CHAPTER 3

HEAVY TAIL DISTRIBUTIONS

The tails of the Gaussian distribution are well understood and are used as a reference for the heaviness of the tails of other distributions. The tails of the Gaussian distribution decay as $\exp(-x^2)$ and a distribution is said to have heavy tails if the tails decays slower than that. The effect of heavy tails is that a random variable with a heavy tail distribution is more likely than a Gaussian variable to exhibit observations far away from the mode of the distribution. A consequence of heavy tails is that the existence of moments is not assured and distributions without second order moment are referred to as infinite variance models. The motivation for considering heavy tail distribution in this thesis is that many real world signals turns out to have heavy tail distributions.

This chapter introduces several heavy tail distributions with focus on application of heavy tail distribution in signal processing. Section 3.1 introduces the class of stable distributions and Section 3.2 introduces other heavy tail distributions. In Section 3.3 estimation in heavy tail distributions and processes, with emphasis on the stable distributions, is discussed.

3.1 Stable Distributions

Stable distributions are a class of probability distributions that allow skewness and heavy tails. There are two reasons why stable distributions are interesting in signal processing. The first is that many real world signals exhibit heavy tails and skewness. The second is the *Generalized Central Limit Theorem* which states that if sums of independent variables converge to a certain distribution, then this distribution is a stable distribution. Thus, the only limit in distribution of normalized sums of independent identical distributed terms is a stable distribution. Additionally to the first empirical argument, some signals may be considered as the sum of many independent identical distributed events, which may justify the stable model as an approximation.

This section establishes the necessary theory on stable distributions for the succeeding part of the thesis. For a thorough introduction on stable distribution see [84, 76] and for a signal processing approach see [73].

3.1.1 Stable Random Variables

An important property of Gaussian random variables is that the sum of two Gaussian random variables is itself a Gaussian random variable. One consequence of this is that if x is Gaussian and x_1, x_2 are iid. Gaussians, then

$$ax_1 + bx_2 \stackrel{d}{=} cx + d$$

where $a, b, c, d \in \mathbb{R}$ and $\stackrel{d}{=}$ means equality in distribution. Let $x \sim N(\mu, \sigma^2)$ then the terms on the left hand side are $N(a\mu, (a\sigma)^2)$ and $N(b\mu, (b\sigma)^2)$, and the term on the right hand side is $N(c\mu + d, (c\sigma)^2)$. Applying the addition rule for independent random variables we have $c^2 = a^2 + b^2$ and $d = (a+b-c)\mu$. Thus the distribution of x is invariant, up to a scale and a shift, under addition. Stable distributions are a class of distributions with this property, and *stable* refers to the fact that the density function is a closed class under addition.

In the literature stable distributions are defined in several equivalent ways. Continuing the considerations above, the following definition is used [84]

DEFINITION 3.1 *A random variable x has a stable distribution if for any positive number a and b, there is a positive number c and a real number d such that*

$$ax_1 + bx_2 \stackrel{d}{=} cx + d$$

where x_1 and x_2 are independent copies of x. If $d = 0$ the random variable is strictly stable.

The pdf. of a sum of two random variables is the convolution between the pdf.s of the random variables, and from Definition 3.1 it follows that the stable distributions are closed with respect to convolution. Thus, a density function, $f(x)$, is stable if for positive numbers b_1, b_2 and real numbers c_1, c_2 there exists a positive number b and a real number c for which

$$f\left(\frac{x-c_1}{b_1}\right) * f\left(\frac{x-c_2}{b_2}\right) = f\left(\frac{x-c}{b}\right)$$

where $*$ denotes convolution [65].

Parameterization Definition 3.1 defines the stable distribution based on the stability property, but the definition does not give a concrete way to parameterize the stable distributions. The most convenient parameterization of stable distributions is through the characteristic function.

PROPOSITION 3.1 *A univariate distribution function is stable if and only if the characteristic function has the form*

$$\boldsymbol{\phi}_x(t) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left(1 - j\beta \operatorname{sign}(t) \tan\frac{\alpha\pi}{2}\right) + j\mu t\right) & , \quad \alpha \neq 1 \\ \exp\left(-\gamma |t|^\alpha \left(1 + j\frac{2}{\pi}\beta \operatorname{sign}(t) \ln|t|\right) + j\mu t\right) & , \quad \alpha = 1 \end{cases} \tag{3.1}$$

where the real parameters α, γ, β, μ satisfy $0 < \alpha \leq 2$, $\gamma > 0$, and $|\beta| \leq 1$.

The stable characteristic function is determined by the four parameters α, γ, β and μ. The parameter α is called the characteristic exponent; the smaller the α-value the more probability mass in the tails of the density function. γ is the scale parameter, β is the symmetry parameter and μ is the location parameter[1].

The characteristic exponent and the symmetry parameter determine the shape of the corresponding density function. In Figure 3.1 the density function for 16 combinations of the *shape* parameters are shown.

[1]There are multiple parameterizations of the αS characteristic function in the literature. In general the characteristic exponent and the symmetry parameter (except for a sign) have the same values, while the scaling and location parameters have different values. For a good discussion of the properties for different parameterizations see [76].

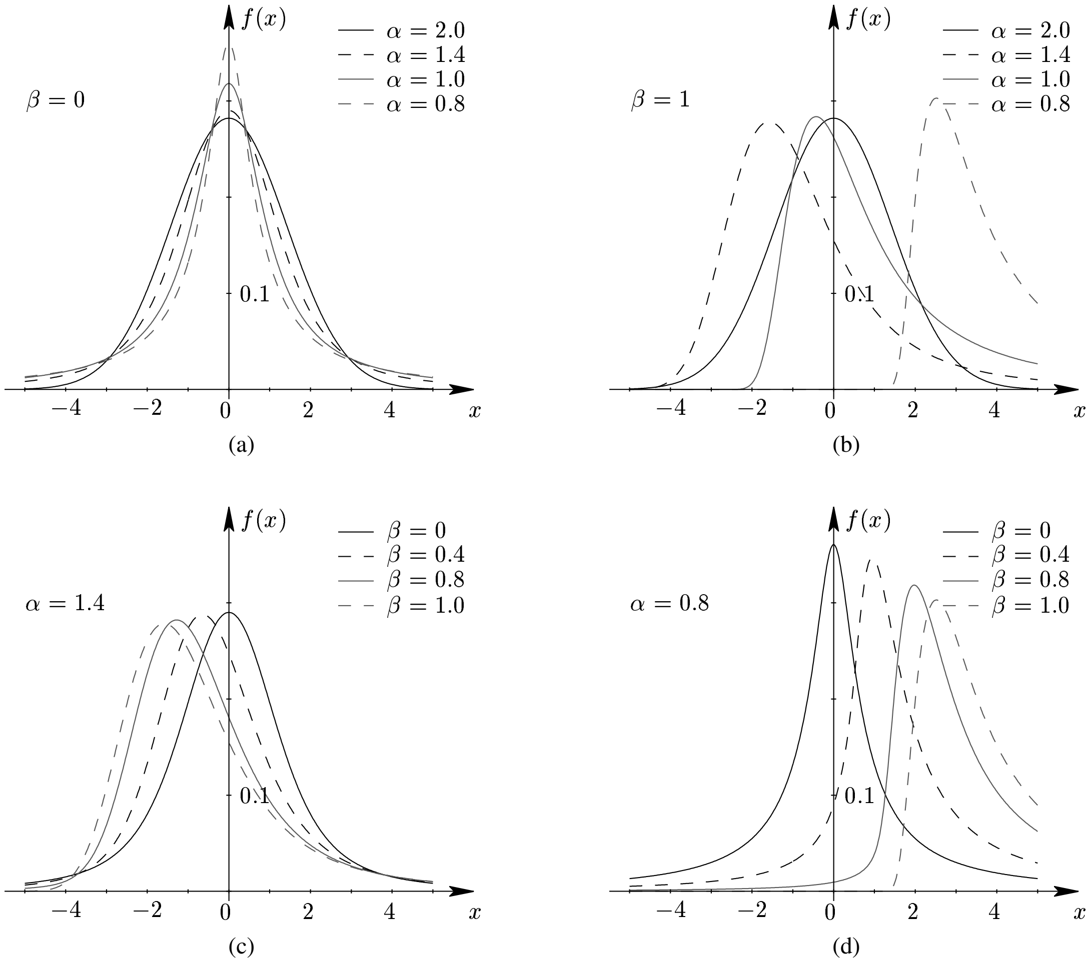


Figure 3.1 Probability distribution functions of $\alpha\mathrm{S}(\alpha, \beta, 1, 0)$ distribution for different values of the characteristic exponent α, and the symmetry parameter β.

To follow the engineering literature stable distribution is referred to as αS , and a random variable x following an αS distribution is denoted $x \sim \alpha\mathrm{S}(\alpha, \beta, \gamma, \mu)$. For $\beta = 0$ and $\mu = 0$ the distribution is symmetric and the characteristic function reduces to

$$\boldsymbol{\phi}_x(t) = \exp\left(-\gamma^\alpha |t|^\alpha\right)$$

Symmetric αS distributions are an important special case and have their own notation: a random variable x following a symmetric αS distribution is denoted $x \sim \mathrm{S}\alpha\mathrm{S}(\alpha, \gamma) = \alpha\mathrm{S}(\alpha, 0, \gamma, 0)$.

The Stable Density Function Explicit expressions for αS density functions are known only in relatively few isolated cases. The lack of expressions for the density function is probably one of the main reasons why αS distribution has received relatively little attention in the field of signal processing. The lack of expressions for the density function is indeed a drawback, but this is to some extent compensated by the progression in numerical computational power.

Three explicit forms of αS distributions are the Gaussian, the Cauchy and the Lévy distributions For $\alpha = 2, \beta = 0$ the αS distribution is a Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

with standard deviation $\sigma = \sqrt{2}\gamma$. For $\alpha = 1, \beta = 0$ the αS distribution is a Cauchy distribution

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\mu)^2}$$

The stable density functions with characteristic exponent $0 < \alpha < 1$ and symmetry parameter $|\beta| = 1$ are bounded functions; for $\beta = 1$ the density is bounded to the right, and for $\beta = -1$ the density is bounded to the left [65]. For $\alpha = 1/2, \beta = -1$ the αS distribution is a Lévy distribution

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\mu)}\right) \quad , \quad x > \mu$$

Even though αS distributions in general can not be expressed on a closed form, the stable density functions can be expanded into absolutely convergent series, see [37].

Properties of αS Random Variables The basic property of αS variables is that sums of αS variables, with identical characteristic exponent is αS

PROPERTY 3.1 *If $u_1 \sim \alpha S(\alpha, \beta_1, \gamma_1, \mu_1)$ and $u_2 \sim \alpha S(\alpha, \beta_2, \gamma_2, \mu_2)$ are independent, then $u_1 + u_2 \sim \alpha S((\alpha, \beta, \gamma, \mu)$ where*

$$\beta = \frac{\beta_1 \gamma_1^\alpha + \beta_2 \gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha} \quad , \quad \gamma = (\gamma_1^\alpha + \gamma_2^\alpha)^{\frac{1}{\alpha}} \quad , \quad \mu = \mu_1 + \mu_2$$

Multiplication and addition with scalars have the following influence on the distribution

PROPERTY 3.2 *If $x \sim \alpha S(\alpha, \beta, \gamma, \mu)$ then for any $a \neq 0, b \in \mathbb{R}$*

$$ax + b \sim \begin{cases} \alpha S(\alpha, \operatorname{sign}(a)\beta, |a|\gamma, a\mu + b) & , \quad \alpha \neq 1 \\ \alpha S(1, \operatorname{sign}(a)\beta, |a|\gamma, a\mu + b - \beta\gamma\frac{2}{\pi}a \ln |a|) & , \quad \alpha = 1 \end{cases}$$

Observe that for $\alpha = 1$ the location parameter is non-linear to a scaling, this is one of the disadvantages of the parameterization in Proposition 3.1 [84]. For $\mu = 0$ we have the symmetry property

PROPERTY 3.3 *For any $0 < \alpha < 2$*

$$x \sim \alpha S(\alpha, \beta, \gamma, 0) \iff -x \sim \alpha S(\alpha, -\beta, \gamma, 0)$$

Even though there are no explicit forms of the general αS pdf.s, certain theoretical properties are known: αS distributions are continuous, and the densities are unimodal and infinitely differentiable. In the Gaussian case the behavior of the tails of the density function is $\exp(-\frac{x^2}{2\sigma^2})$, whereas for $\alpha < 2$ we have the tail property [84]

PROPERTY 3.4 *If $v \sim \alpha S(\alpha, \beta, \gamma, \mu)$ with $0 < \alpha < 2$, $-1 < \beta < 1$ then*

$$\lim_{x\to\infty} f_v(x) = C_\alpha \frac{1-\beta}{2} \alpha\gamma^\alpha x^{-(\alpha+1)} \quad \textit{where} \quad C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \textit{if } \alpha \neq 1 \\ \frac{2}{\pi} & \textit{if } \alpha = 1 \end{cases} \tag{3.2}$$

Using the symmetry property, the lower tail properties are similar.

Thus the asymptotic tail behavior is algebraic (inverse power law). The tail properties of the SαS density function are shown in Figure 3.2

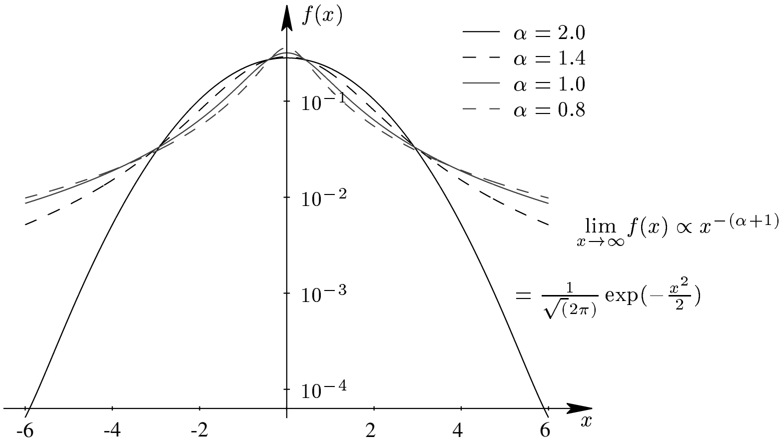


Figure 3.2 Probability density functions of $\mathrm{S}\alpha\mathrm{S}(\alpha, 1)$ depicted on a logarithmic scale to show the behavior of the tails.

Domain of Attraction The classical Central Limit Theorem state that the normalized sum of iid. random variables with finite second order moment converges towards a Gaussian distribution. Thus any distribution with finite second order moment is *attracted* by the Gaussian distribution, and all such distributions are said to be in the *Domain of Attraction* of the Gaussian distribution.

The Generalized Central Limit Theorem does not have restrictions on the first and second order moment [76]

THEOREM 3.1 *Generalized Central Limit Theorem. Let $u_1, u_2, \ldots, u_N$ be iid. random variables. There exist constants $a > 0, b \in \mathbb{R}$*

$$a(u_1 + \cdots + u_N) - b \xrightarrow{d} v$$

if and only if v is αS , in which case $a = N^{-1/\alpha}$ for some $0 < \alpha \leq 2$

The notation $\xrightarrow{d}$ denote convergence in distribution. From the Generalized Central Limit Theorem it appears that the only pdf.s that have a non-empty domain of attraction are the αS distributions [37].

Generating αS Variables In simulation studies there is a need for generating synthetic signals. One way to generate an arbitrary random variable is to generate a uniformly distributed variable and transform this variable with the inverse of the desired distribution. However, determining the inverse distribution might be difficult or impossible, which is the case for αS distributions in general. Furthermore the method is very sensitive to inaccuracies in the inverse distribution.

A few specific αS variables are easy to generate. Based on a random variable v, uniform on the interval $[-\pi/2\ ;\ \pi/2]$, the random variable $x = \gamma \tan(v)$ is Cauchy distributed, i.e. $x \sim \mathrm{S}\alpha\mathrm{S}(1, \gamma)$. To generate a Lévy distributed variable, $x \sim \alpha\mathrm{S}(\frac{1}{2}, 1, \gamma, \mu)$, let $x = (\gamma/u)^2 + \mu$ where $u \sim N(0, 1)$. It is also easy to generate an $\mathrm{S}\alpha\mathrm{S}(\alpha, 1)$ distributed variable; let u be an exponential distributed variable with $\mathrm{E}\,(u) = 1$ and v a uniform distributed variable on $[-\pi/2\ ,\ \pi/2]$, then

$$x = \frac{\sin(\alpha v)}{\cos(v)^{1/\alpha}} \left(\frac{\cos((1-\alpha)v)}{u} \right)^{\frac{1-\alpha}{\alpha}}$$

is $\alpha\mathrm{S}(\alpha, 0, 1, 0)$ [84]. In the Gaussian case, $\alpha = 2$, the equation reduces to $x = \sin(2v)\cos(v)^{-1}u^{1/2} = 2\sin(v)u^{1/2}$, which is the Box-Muller method for generating $N(0, 2)$ random variables [15].

The generation of general αS variables is in general more complicated. First consider the generation of $\alpha\mathrm{S}(\alpha, 1, 1, 0)$ variables. Again let u be a standard exponential distributed variable and v an uniform distributed variable on $[-\pi/2\ ,\ \pi/2]$, then (for $\alpha \neq 1$)

$$x = \sin((1-\alpha)v)^{\frac{1-\alpha}{\alpha}} \frac{\sin(\alpha v)}{\sin(v)} u^{-\frac{1-\alpha}{\alpha}}$$

is $x \sim \alpha\mathrm{S}(\alpha, 1, 1, 0)$. Now return to the general $\alpha\mathrm{S}(\alpha, \beta, \gamma, 0)$ case. One method is based on a transformation of 2 iid. $\alpha\mathrm{S}(\alpha, 1, \gamma, 0)$ variables. Let v_1, v_2 be two independent $\alpha\mathrm{S}(\alpha, 1, \gamma, 0)$, and let

$$x = \begin{cases} \left(\frac{1+\beta}{2}\right)^{1/\alpha} v_1 - \left(\frac{1-\beta}{2}\right)^{1/\alpha} v_2 & \text{if } \alpha \neq 1 \\ \left(\frac{1+\beta}{2}\right) v_1 - \left(\frac{1-\beta}{2}\right) v_2 + \gamma^{1/\alpha}\left(\frac{1+\beta}{\pi}\ln\frac{1+\beta}{2} - \frac{1-\beta}{\pi}\ln\frac{1-\beta}{2}\right) & \text{if } \alpha = 1 \end{cases}$$

then $x \sim \alpha\mathrm{S}(\alpha, \beta, \gamma, 0)$ [84]. Throughout this work an algorithmic version of this method, proposed in [26], has been applied for the generation of αS random variables.

3.1.2 Stable Random Vectors

The multivariate αS distribution is defined by extending the definition of an αS random variable to $\mathbb{R}^d$ [84]

DEFINITION 3.2 *A random vector $\mathbf{u} = [u_1, u_2, \ldots, u_d]^T$ is said to be a stable random vector in $\mathbb{R}^d$, if for any positive number a and b there is a positive number c such that*

$$a\mathbf{u}_1 + b\mathbf{u}_2 \stackrel{d}{=} c\mathbf{u}$$

where $\mathbf{u}_1$ and $\mathbf{u}_2$ are independent and follow the distribution of $\mathbf{u}$.

That is, a random vector is an αS vector if and only if all of its components are αS random variables, and the marginal distribution are all αS with the same characteristic exponent.

The joint characteristic function of an αS random vector $\mathbf{u}$ in $\mathbb{R}^d$ is an αS characteristic function on the form in Theorem 3.2.

THEOREM 3.2 *A d-dimensional distribution function, $f(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^d$, is stable if and only if the corresponding characteristic function has the form*

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{t}) = \begin{cases} \exp\left(-\int_{S_d} |\mathbf{t}^T\mathbf{s}|^\alpha \left(1 - j\,\mathrm{sign}(\mathbf{t}^T\mathbf{s})\tan\frac{\alpha\pi}{2}\right)\Gamma(d\mathbf{s}) + j\mathbf{t}^T\boldsymbol{\mu}_{\mathbf{x}}\right) & , \quad \alpha \neq 1 \\ \exp\left(-\int_{S_d} |\mathbf{t}^T\mathbf{s}|^\alpha \left(1 + j\frac{2}{\pi}\mathrm{sign}(\mathbf{t}^T\mathbf{s})\ln|\mathbf{t}^T\mathbf{s}|\right)\Gamma(d\mathbf{s}) + j\mathbf{t}^T\boldsymbol{\mu}_{\mathbf{x}}\right) & , \quad \alpha = 1 \end{cases}$$

$\boldsymbol{\mu}, \mathbf{t} \in \mathbb{R}^d$. S_d is the d-dimensional unit sphere, $S_d = \{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\| = 1\}$, $\Gamma(d\mathbf{s})$ is a finite measure on S_d, and α is the characteristic exponent. The pair $(\Gamma(\mathbf{s}), \boldsymbol{\mu}_{\mathbf{x}})$ is unique.

The expression for the characteristic function involves an integration over the unit sphere S_d. The measure Γ is called the spectral measure and the vector $\boldsymbol{\mu}_{\mathbf{x}}$ is called the location vector of the αS random vector. Thus, while the univariate characteristic function forms a parametric set, the multivariate characteristic function in general forms a non-parametric set.

Symmetric Stable Distributions As for the univariate αS distribution the SαS vector is an important special case. A d-dimensional pdf. $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ with $0 < \alpha < 2$, is SαS if the corresponding chf. has the form

$$\boldsymbol{\phi}_\alpha(\mathbf{t}) = \exp\left(-\int_{S_d} |\mathbf{t}^T\mathbf{s}|^\alpha \Gamma(d\mathbf{s})\right)$$

where $\mathbf{t} \in \mathbb{R}^d$, $S_d = \{\mathbf{s} \in \mathbb{R}^d : \|\mathbf{s}\| = 1\}$, and $\Gamma(\mathbf{s})$ is the spectral measure on S_d

Multivariate Gaussian Distribution For $\alpha = 2$ the αS vector is the multivariate Gaussian distribution with the well-known chf.

$$\boldsymbol{\phi}(\mathbf{t}) = \exp\left(-\mathbf{t}^T\mathbf{A}\mathbf{t} + j\mathbf{t}^T\boldsymbol{\mu}\right)$$

with mean vector $\boldsymbol{\mu}$ and covariance matrix $2\mathbf{A}$.

Stable Random Vector with Discrete Spectral Measure Consider the random vector $\mathbf{v}$ with independent elements $v_n \sim \alpha\mathrm{S}(\alpha, \beta_n, \gamma_n, \mu_n)$, $n = 1, \ldots, N$, and let $\mathbf{A}$ be a $M{\times}N$ matrix. The chf. of the random vector $\mathbf{x} = \mathbf{A}\mathbf{v}$ is

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{t}) = \mathrm{E}\exp\left(j\mathbf{t}^T\mathbf{x}\right) = \mathrm{E}\exp\left(j\mathbf{t}^T\mathbf{A}\mathbf{v}\right) = \mathrm{E}\exp\left(j\sum_{n=1}^{N}\mathbf{t}^T\mathbf{a}_{\cdot n}v_n\right) \tag{3.3}$$

Since the elements of $\mathbf{v}$ are independent, the ch.f. can be written as

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{t}) = \prod_{n=1}^{N}\mathrm{E}\exp\left(j\mathbf{t}^T\mathbf{a}_{\cdot n}v_n\right) = \prod_{n=1}^{N}\phi_{v_n}(\mathbf{t}^T\mathbf{a}_{\cdot n}) \tag{3.4}$$

From Definition 3.2 it follows that $\mathbf{x}$ is a stable vector. Using Proposition 3.1 and Equation 3.4 the characteristic function is

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{t}) = \begin{cases} \exp\sum_{n=1}^{N}\left(-\gamma_n^\alpha|\mathbf{t}^T\mathbf{a}_{\cdot n}|^\alpha\left(1 - j\beta_n\mathrm{sign}(\mathbf{t}^T\mathbf{a}_{\cdot n})\tan\frac{\alpha\pi}{2}\right) + j\mu_n\mathbf{t}^T\mathbf{a}_{\cdot n}\right) & , \quad \alpha \neq 1 \\ \exp\sum_{n=1}^{N}\left(-\gamma_n^\alpha|\mathbf{t}^T\mathbf{a}_{\cdot n}|^\alpha\left(1 + j\beta_n\mathrm{sign}(\mathbf{t}^T\mathbf{a}_{\cdot n})\frac{2}{\pi}\ln|\mathbf{t}^T\mathbf{a}_{\cdot n}|\right) + j\mu_n\mathbf{t}^T\mathbf{a}_{\cdot n}\right) & , \quad \alpha = 1 \end{cases}$$

Defining $\mathbf{s}_n = \mathbf{a}_{\cdot n}/\|\mathbf{a}_{\cdot n}\|_2$ the characteristic function is

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{t}) = \begin{cases} \exp\Bigg(-\sum_{n=1}^{N}\gamma_n^\alpha|\mathbf{t}^T\mathbf{s}_n|^\alpha\|\mathbf{a}_{\cdot n}\|_2^\alpha\frac{1+\beta_n}{2}\left(1 - j\mathrm{sign}(\mathbf{t}^T\mathbf{a}_{\cdot n})\tan\frac{\alpha\pi}{2}\right) \\ \quad -\sum_{n=1}^{N}\gamma_n^\alpha|\mathbf{t}^T\mathbf{s}_n|^\alpha\|\mathbf{a}_{\cdot n}\|_2^\alpha\frac{1-\beta_n}{2}\left(1 + j\mathrm{sign}(\mathbf{t}^T\mathbf{a}_{\cdot n})\tan\frac{\alpha\pi}{2}\right) \\ \quad +\sum_{n=1}^{N}j\mu_n\mathbf{t}^T\mathbf{a}_{\cdot n}\Bigg) & , \quad \alpha \neq 1 \\ \exp\Bigg(-\sum_{n=1}^{N}\gamma_n|\mathbf{t}^T\mathbf{s}_n|^\alpha\|\mathbf{a}_{\cdot n}\|_2^\alpha\frac{1+\beta_n}{2}\left(1 + j\frac{2}{\pi}\mathrm{sign}(\mathbf{t}^T\mathbf{a}_{\cdot n})\ln|\mathbf{t}^T\mathbf{a}_{\cdot n}|\right) \\ \quad -\sum_{n=1}^{N}\gamma_n|\mathbf{t}^T\mathbf{s}_n|^\alpha\|\mathbf{a}_{\cdot n}\|_2^\alpha\frac{1-\beta_n}{2}\left(1 - j\frac{2}{\pi}\mathrm{sign}(\mathbf{t}^T\mathbf{a}_{\cdot n})\ln|\mathbf{t}^T\mathbf{a}_{\cdot n}|\right) \\ \quad +\sum_{n=1}^{N}j\mu_n\mathbf{t}^T\mathbf{a}_{\cdot n}\Bigg) & , \quad \alpha = 1 \end{cases}$$

Rewriting to the form of the characteristic function in Theorem 3.2

$$\boldsymbol{\phi}_{\mathbf{x}}(\mathbf{t}) = \begin{cases} \exp\left(-\int_{S_d} |\mathbf{t}^T\mathbf{s}_n|^\alpha \left(1 - j\,\mathrm{sign}(\mathbf{t}^T\mathbf{s}_n)\tan\frac{\alpha\pi}{2}\right)\Gamma(d\mathbf{s}) + j\mathbf{t}^T\boldsymbol{\mu}_{\mathbf{x}}\right) & , \quad \alpha \neq 1 \\ \exp\left(-\int_{S_d} |\mathbf{t}^T\mathbf{s}_n|^\alpha \left(1 + j\frac{2}{\pi}\mathrm{sign}(\mathbf{t}^T\mathbf{s}_n)\ln|\mathbf{t}^T\mathbf{s}_n|\right)\Gamma(d\mathbf{s}) + j\mathbf{t}^T\boldsymbol{\mu}_{\mathbf{x}}\right) & , \quad \alpha = 1 \end{cases}$$

where

$$\Gamma(\mathbf{s}) = \sum_{n=1}^{N} \gamma_n^\alpha \|\mathbf{a}_{\cdot n}\|_2^\alpha \frac{1+\beta_n}{2}\delta(\mathbf{s}-\mathbf{s}_n) + \sum_{n=1}^{N} \gamma_n^\alpha \|\mathbf{a}_{\cdot n}\|_2^\alpha \frac{1-\beta_n}{2}\delta(\mathbf{s}+\mathbf{s}_n) \tag{3.5}$$

and

$$\boldsymbol{\mu}_{\mathbf{x}} = \begin{cases} \sum_{n=1}^{N} j\mu_n \mathbf{a}_{\cdot n} & , \quad \alpha \neq 1 \\ \sum_{n=1}^{N} j\left(\mu_n \mathbf{a}_{\cdot n} - \frac{2}{\pi}\beta_n \ln\left(\mathbf{a}_{\cdot n}^T \mathbf{a}_{\cdot n}\right)\right) & , \quad \alpha = 1 \end{cases}$$

Equation 3.5 shows that the spectral measure Γ of the random vector $\mathbf{x}$ is discrete with masses on N symmetric pairs of points of S_d. Conversely, an αS random vector with discrete spectral measure can always be represented as a linear transformation of independent αS variables. Thus

PROPOSITION 3.2 *The spectral measure of an αS vector is concentrated on a finite number of points on the unit sphere if and only if the vector can be expressed as a linear transformation of independent αS random variables.*

Example: 2-dimensional αS densities The contour curves for a 2-dimensional Gaussian pdf. are elliptical in all levels. Elliptical contours are possible for non-Gaussian αS pdf.s if and only if the spectral measure is continuous and distributed in a precise way (see [84]). For non-Gaussian αS pdf.s with discrete spectral measure the levels are convex near the mode of the pdf. but becomes non-convex away from the mode.

In Figure 3.3 the contours for four different 2-dimensional αS pdf.s are shown. The corresponding spectral measure is discrete and can be written in the form

$$\Gamma(\mathbf{s}) = \sum_{n=1}^{N} \gamma_n \delta(\mathbf{s}-\mathbf{s}_n) \tag{3.6}$$

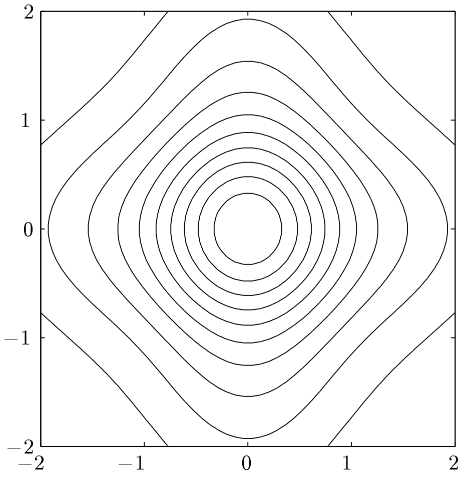
where $\delta(\mathbf{s}-\mathbf{s}_n)$ is a delta function assigning unity mass to the direction $\mathbf{s}_n = [\cos(\theta_n)\ \sin(\theta_n)]^T$. The characteristic exponent, the directions and the values of the spectral measure for the specific directions are given in the tables above the figures.

Figure 3.3(a) shows the pdf. contour curves for a SαS vector with iid. elements and characteristic exponent $\alpha = 1.0$, this is a 2-dimensional Cauchy distribution. The pdf.s in Figure 3.3(b) and 3.3(c) have identical spectral measure with characteristic exponent $\alpha = 1.0$ and $\alpha = 0.8$ respectively. It appears that the lower the α-value the closer the non-convex contour level curves come to the mode.

The contour curves in Figure 3.3(b), 3.3(c) and 3.3(d) can be considered as the contours of the pdf. for a vector $\mathbf{x}$, where $\mathbf{x} = \mathbf{A}\mathbf{v}$, $\mathbf{v} \in \mathbb{R}^3$ is a vector with iid. αS elements and

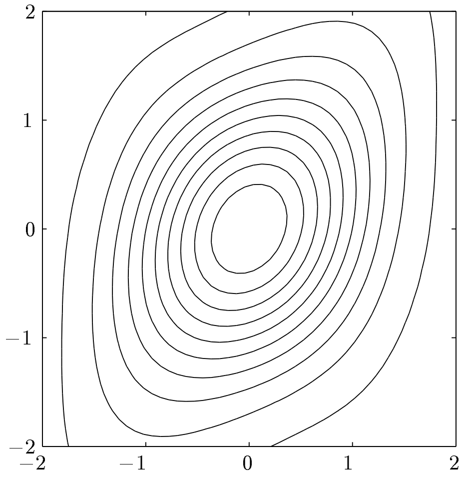
$$\mathbf{A} = \begin{bmatrix} \cos(0) & \cos(2\pi/3) & \cos(4\pi/3) \\ \sin(0) & \sin(2\pi/3) & \sin(4\pi/3) \end{bmatrix} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$$

$\alpha = 1.0$	$n = 1$	$\theta_n = \quad 0$	$\gamma_n = 1/2$
	2	$\pi/2$	$1/2$
	3	π	$1/2$
	4	$3\pi/2$	$1/2$



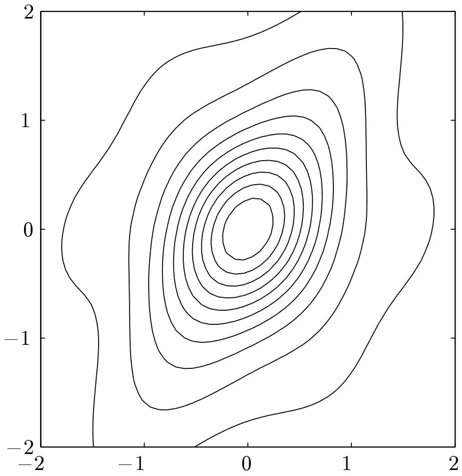
(a)

$\alpha = 1.0$	$n = 1$	$\theta_n = \quad 0$	$\gamma_n = 1/4$
	2	$\pi/3$	$1/2$
	3	$2\pi/3$	$1/4$
	4	π	$1/4$
	5	$4\pi/3$	$1/2$
	6	$5\pi/3$	$1/4$



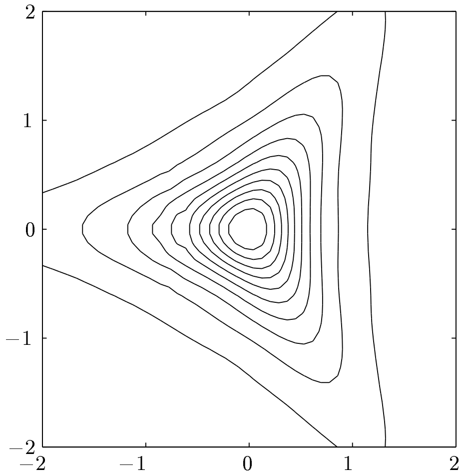
(b)

$\alpha = 0.8$	$n = 1$	$\theta_n = \quad 0$	$\gamma_n = 1/4$
	2	$\pi/3$	$1/2$
	3	$2\pi/3$	$1/4$
	4	π	$1/4$
	5	$4\pi/3$	$1/2$
	6	$5\pi/3$	$1/4$



(c)

$\alpha = 0.8$	$n = 1$	$\theta_n = \quad 0$	$\gamma_n = 1/2$
	2	$2\pi/3$	$1/2$
	3	$4\pi/3$	$1/2$



(d)

Figure 3.3 Contours for four different 2-dimensional αS density functions. The spectral measure corresponding to the densities is given on the form in Equation 3.6 and the tables above the contour plots list the parameters. For calculation of multidimensional αS densities see [75].

The different pdf.s are shaped by the pdf. of the vector $\mathbf{v}$. The random vector $\mathbf{v} = [v_1\ v_2\ v_3]^T$ corresponding to the pdf. in Figure 3.3(b), 3.3(c) and 3.3(d) are respectively $v_1 \sim \mathrm{S}\alpha\mathrm{S}(1, 1/4)$, $v_2 \sim \mathrm{S}\alpha\mathrm{S}(1, 1/4)$, $v_3 \sim \mathrm{S}\alpha\mathrm{S}(1, 1/2)$; $v_1 \sim \mathrm{S}\alpha\mathrm{S}(0.8, 1/4)$, $v_2 \sim \mathrm{S}\alpha\mathrm{S}(0.8, 1/4)$, $v_3 \sim \mathrm{S}\alpha\mathrm{S}(0.8, 1/2)$, and $v_1 \sim \alpha\mathrm{S}(0.8, -1, 1/2, 0)$, $v_2 \sim \alpha\mathrm{S}(0.8, -1, 1/2, 0)$, $v_3 \sim \alpha\mathrm{S}(0.8, -1, 1/2, 0)$. In the scope of the thesis it is natural to think of $\mathbf{x}$ to be the observable signal vector from an instantaneous mixture of 3 independent sources, where $\mathbf{v}$ is the source signal vector.

Sub-Gaussian Random Vector An important subclass of multivariate SαS random vectors are the sub-Gaussian random vectors[2]. If u is a positive $\alpha\mathrm{S}(\alpha/2, 1, \gamma, \mu)$ random variable and v is a $\mathrm{N}(0, \sigma^2)$ random variable, then the random variable $x = u^{1/2}v$ is $\mathrm{S}\alpha\mathrm{S}(\alpha, \gamma)$ [84]. This result extends to a random vector $\mathbf{x}$. Let

$$u \sim \alpha\mathrm{S}(\alpha, 1, \cos(\frac{\pi\alpha}{2})^{2/\alpha}, 0)$$

and $\mathbf{v} \in \mathbb{R}^{\mathbf{d}}$ be a random zero mean Gaussian vector independent of u, then $\mathbf{x} = u^{1/2}\mathbf{v}$ is a d-dimensional SαS random vector. The vector $\mathbf{x}$ is said to be a sub-Gaussian vector, and it has chf. on the form

$$\boldsymbol{\phi}_\alpha(\mathbf{t}) = \exp\left(-\left|\frac{1}{2}\mathbf{t}^T\mathbf{R}\mathbf{t}\right|^{\alpha/2}\right)$$

where $0 < \alpha < 2$, $\mathbf{t} \in \mathbb{R}^d$, and the matrix $\mathbf{R}$ is positive definite [84]. The sub-Gaussian random vector $\mathbf{x}$ is denoted $x \sim \mathrm{SG}(\mathbf{R})$ where $\mathbf{R}$ is the covariance matrix of the random vector $\mathbf{v}$. A sub-Gaussian vector has a continuous spectral measure, i.e. it can not be considered as a linear combination of a finite number of iid. components.

3.2 Other Heavy Tail Distributions and Heavy Tail Processes

In Section 3.1 attention is on the class of stable distributions. This section introduces another class of heavy tail distributions termed the Normal Inverse Gaussian distribution, and a few other heavy tail distributions are briefly mentioned. Furthermore the theory of stochastic processes introduced in Section 2.3.1 are reviewed for the case of heavy tail distributed excitation processes.

3.2.1 Normal Inverse Gaussian Distribution

The Normal Inverse Gaussian (NIG) distribution is a versatile non-Gaussian distribution and it belongs to the class of Generalized Hyperbolic distributions. The pioneering work on NIG distributions was done by Barndorff-Nielsen in 1977 and 1995, more recent references are [8, 9].

The NIG distribution has characteristics that make it an interesting distribution for modelling signals with heavy tail pdf.s, and some of the properties of the NIG distribution are similar to the properties for αS distributions. Contrary to the αS distribution, the NIG has an explicit form of the density function.

DEFINITION 3.3 *A random variable v is NIG if it has probability density function on the form*

$$f_v(x) = \frac{\alpha\delta}{\pi}\exp(\delta\sqrt{\alpha^2 - \beta^2} - \beta\mu)\frac{K_1(\alpha\sqrt{\delta^2 + x^2})}{\sqrt{\delta^2 + x^2}}\exp(\beta x)$$

where μ, $\delta > 0$, $0 \le |\beta| \le \alpha \in \mathbb{R}$ and K_1 is the modified Bessel function of the second kind and index 1.

[2]Notice: The term *Sub-Gaussian* is sometimes used to refer to heavy tail distribution in general. This practice is not used here since there is a specific distribution that has the name sub-Gaussian.

A random variable v following a NIG pdf. is denoted $v \sim \mathrm{NIG}(\alpha, \beta, \delta, \mu)$. The NIG distribution is parameterized by four parameters α, β, δ and μ, and the parameters have an interpretation similar to the parameters for the αS distribution: α determines the tail behavior, the lower α the heavier the tails; β is the symmetry parameter, $\beta = 0$ gives a symmetric density, $\beta > 0$ implies a density skew to the right and $\beta < 0$ implies a density skew to the left; δ is the scale parameter; and μ is the location parameter. To get an impression of the versatility of the NIG distribution the pdf.s for different values of the α and β parameters are depicted in Figure 3.4.

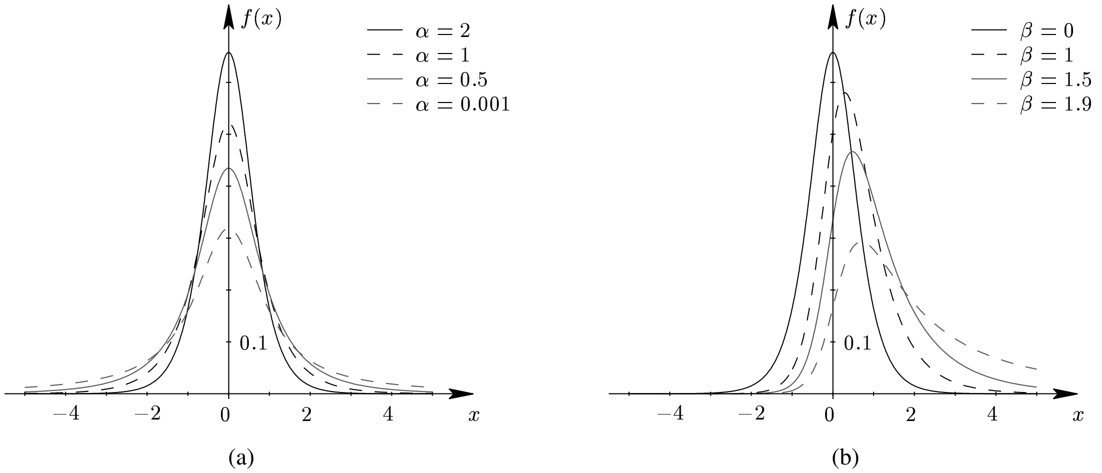


Figure 3.4 Plot of NIG probability density functions for different values of the α and β parameter. (a) is $NIG(\alpha, 0, 1, 0)$ densities for $\alpha = 0.001, 0.5, 1, 2$. (b) is $NIG(2, \beta, 1, 0)$ densities for $\beta = 0, 1, 1.5, 1.9$.

A method for synthetically generation of iid. NIG random variables, based on a mixture of a Gaussian and an Inverse Gaussian variable, is proposed in [82]. The Inverse Gaussian pdf. is

$$f(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{-\frac{1}{2}} \exp\left(-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right) \tag{3.7}$$

and an Inverse Gaussian variable parameterized by (μ, λ) is denoted $u \sim IG(\mu, \lambda)$. A NIG random variable $x \sim NIG(\alpha, \beta, \delta, \mu)$ can be generated as $x = \mu + \beta u + v\sqrt{u}$, where $u \sim IG(\delta^2, \alpha^2 - \beta^2)$ and $v \sim N(0, 1)$.

For fixed values of α, δ and μ the class of NIG distributions establishes an exponential model with β as canonical parameter. For $v \sim \mathrm{NIG}(\alpha, \beta, \delta, \mu)$ the chf. is

$$\phi_v(t) = \exp\left(\delta\sqrt{\alpha^2 - \beta^2} - \delta\sqrt{\alpha^2 - (\beta + jt)^2} + j\mu t\right) \tag{3.8}$$

[8]. From the chf. it appears that the NIG distributions are infinitely divisible. Consequently, if $v_1, \ldots, v_N$ are independent random variables and $v_n \sim \mathrm{NIG}(\alpha, \beta, \delta_n, \mu_n)$ then the sum $x = \sum_{n=1}^{N} v_n$ is also a NIG distributed random variable with $x \sim \mathrm{NIG}(\alpha, \beta, \delta, \mu)$ and $\delta = \sum_{n=1}^{N} \delta_n$ and $\mu = \sum_{n=1}^{N} \mu_n$. This property is similar to the property that defines the class of stable pdf.s, however the NIG pdf. is not stable. One way to see this is that the scale parameter δ diverge for an infinite normalized sum of NIG random variables.

From the chf. in Equation 3.8 it is also seen that the Gaussian and the Cauchy distributions are special cases of the NIG distribution. The Gaussian distribution appears as the limiting case $\beta = 0$, $\alpha \to \infty$ and where $\sigma^2 = \delta/\alpha$, and $\mathrm{NIG}(0, 0, \delta, \mu)$ is the Cauchy distribution [8].

In Section 3.1.1 it is established that tails of the αS pdf.s are algebraic in the asymptote. In [46] the tail behavior of the NIG pdf.s is shown to be

$$\lim_{|x|\to\infty} f(x) \propto \begin{cases} |x|^{-3/2}\exp(\beta x - \alpha|x|) & , \quad \alpha \neq 0 \\ |x|^{-2} & , \quad \alpha \to 0 \end{cases}$$

For $\alpha \neq 0$ the asymptotic tail behavior exhibits a combined algebraic and exponential decay, where the exponential term is determined by the α and β parameters. For $\alpha \to 0$ the NIG approaches the Cauchy distribution and obviously the asymptotic tail behavior is approaching the Cauchy tail.

3.2.2 Related Distributions

It is appropriate to mention a few other heavy tail distributions that have found applications in signal processing. One of the most common heavy tail distributions is the double exponential distribution that is considered in Chapter 2. It appears that it is a heavy tail distribution since the tails in the pdf. decay as $\exp(-|x|)$.

The Pareto pdf. with tail parameter α and scale parameter c is defined as [36]

$$f(x) = \frac{\alpha c^{\alpha}}{x^{1+\alpha}} \quad \text{for} \quad x > c \tag{3.9}$$

It appears that the tail behavior of the Pareto distribution is similar to the tail behavior of the stable distributions. The Pareto distribution function is $F(x) = 1 - (\frac{c}{x})^{\alpha}$, and since there is a closed form expression of the inverse of the distribution function it is easy to generate a Pareto distributed random variable from a uniform random variable.

Another interesting heavy tail distribution is the t-distribution. A random variable is t-distributed if it has pdf. of the form [36]

$$f(x) = \frac{\Gamma(\frac{1}{2}(n+1))}{\sqrt{n\pi}\,\Gamma(\frac{1}{2}n)}\left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)} \tag{3.10}$$

where n is the degree of freedom. A random variable x following a t-distribution with n degrees of freedom is denoted $x \sim t(n)$. For $n = 1$ the t-distribution is equal to the Cauchy distribution, and in the limit $n \to \infty$ the distribution is identical to the Gaussian distribution. An interesting property is that the degree of freedom is the upper bound for the order of moment that exists, i.e. the first moment exists for $n > 1$, the second moment exists for $n > 2$, and so on.

It is interesting to notice that the t-distribution share an interesting property with the αS and NIG distributions, namely that the Gaussian and the Cauchy distributions appear as special cases. That is, in the complete space of distributions described by the three classes of distribution, there are two points where the three classes are identical.

3.2.3 Heavy Tail Stochastic Processes

For stationary stochastic processes two models are widely used. In the time domain a linear model is used, and a stationary random signal is represented by a time invariant linear filter with iid. input signal. In the frequency domain a Fourier transform model is used and a stationary random signal is represented as a superposition of harmonics with random amplitude and phase [80]. For stationary processes with finite second order moment these two models are nearly interchangeable, and for Gaussian processes the two models are equivalent. For infinite second order moment stationary processes, this duality is in general not maintained.

Although both continuous time processes and skewed stable distributions are important in certain applications only discrete time SαS processes are addressed here. A stochastic process u_n, where $n \in T$ and T is an index set, is a SαS process if all finite-dimensional distributions of

$$[u_{n_1}, u_{n_2}, \ldots, u_{n_d}]^T \quad , \quad n_1, n_2, \ldots, n_d \in T$$

are SαS . Two basic types of SαS stochastic processes, that correspond to the time and frequency models, are of special interest here: *Linear Stable Processes* and *Harmonizable Stable processes*.

Linear Stable Processes Let u_n, where $n \in T$ and T is an index set, be iid. SαS random variables. Then, for some appropriate bounds on the coefficients a_l,

$$x_n = \sum_{l=-\infty}^{\infty} a_l u_{n-l}$$

is a linear SαS stationary process.

Harmonizable Stable Processes From the theory of stochastic processes it is known that a discrete-time weakly stationary process, x_n, has a spectral representation on the form

$$x_n = \int_{-\pi}^{\pi} e^{jn\omega} dZ(\omega)$$

where $dZ(\omega)$ is a stochastic process of orthogonal increments. If a stationary SαS process has a spectral representation, under some appropriate interpretation of the integral and some conditions on $dZ(\omega)$, then the process is a harmonizable stable process [19, 80].

For heavy tailed stationary processes with infinite second order moments the interchangeability of time and frequency domian models is not guaranteed, and for SαS processes the classes of linear (Moving Average) and harmonizable stable processes are disjoint classes. That is, harmonizable stable processes can not be represented as linear stable processes and linear stable processes are not harmonizable [19, 84].

3.3 Estimation in Heavy Tail Processes

One consequence of heavy tail distributions is that not all moments exist, thus classical estimation methods must be applied with caution. This section considers estimation in heavy tail distributions and processes. The theory on moments from Section 2.1.2 is reviewed and methods for estimation in uni- and multivariate distributions are considered.

3.3.1 Moments and Heavy Tail Distributions

Often the first and the second moment are routinely used to describe the distribution, however these are in general not useful for heavy tailed distributions, because the integral expressions for these expectations may diverge. In the place of integer moments it is sometimes useful to use fractional moments. Here we introduce fractional moments, discuss the existence of moments, and introduce normalized moments and cumulants.

In Section 2.1.2 the moment generating function and the integer moments are defined. For a random variable u with pdf. $f_u(x)$ the pth absolute moment, $\mathrm{E}(|u|^p)$, can be written as the integral

$$\mathrm{E}(|u|^p) = \int_{-\infty}^{\infty} |x|^p f_u(x) dx \tag{3.11}$$

Thus the first, second and higher order absolute moments are obtained for $p \in \mathbb{Z}^+$, and the fractional moments are obtained for $0 < p, p \in \mathbb{R}$. As mentioned the existence of the pth moment is not assured for all pdf.s for all p, since the existence requires that the integral in Equation 3.11 converges. Clearly the existence of the pth moment is strongly dependent on the tail behavior of the pdf..

For $\alpha \neq 0$ the chf. of the NIG distribution in Equation 3.8 has a Taylor expansion. Thus, even though the NIG distribution is a heavy tail distribution all moments, except in the limit $\alpha \to 0$, exist.

For non-Gaussian αS distributions the second order moment does not exist, however from the tail behavior of the αS pdf. in Equation 3.2 it is easy to show that all fractional moments of order $0 < p < \alpha$ do exist. For SαS distributions we have the following property [73]

PROPERTY 3.5 *The pth order moment of the random variable $u \sim S\alpha S(\alpha, \gamma)$ is*

$$\mathrm{E}\left(|u|^p\right) = C(p,\alpha)\gamma^p \qquad -1 < p < \alpha \tag{3.12}$$

where

$$C(p,\alpha) = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{-p}{\alpha}\right)}{\alpha\sqrt{\pi}\Gamma\left(\frac{-p}{2}\right)}$$

and where $\Gamma(\cdot)$ is the usual Γ-function.

Notice that the SαS pdf. also has finite negative moments. In Figure 3.5 the value of the pth moment for $u \sim \mathrm{S}\alpha\mathrm{S}(\alpha, 1)$ versus the moment order is depicted for different values of the characteristic exponent.

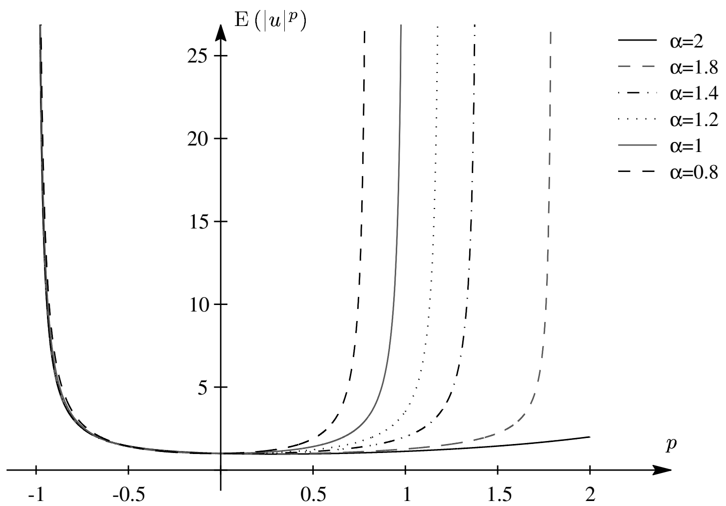


Figure 3.5 The pth order moment of $u \sim \mathrm{S}\alpha\mathrm{S}(\alpha, 1)$ versus moment order for different values of α.

For $0 < p < 2$ the moments are called fractional lower order moments (FLOM) and for $-1 < p < 0$ the moments are called negative order moments (NOM). In the context of standard statistical signal processing most methods are based on second order moments. Methods based on higher order moments (HOM) and recently also methods based on fractional lower order moments have emerged. It may be useful to think of moments as a class of functions applicable to statistical signal processing.

Normalized Moments and Cumulants For non-Gaussian αS random variables the second and higher order moments do not exist. However, for linear SαS processes the normalized moment and correlation functions exist. Consider the linear αS process

$$v(n) = \sum_{k=-\infty}^{\infty} h(k)u(n-k)$$

where $u(n)$ are iid. SαS . Let $r_h(m)$ denote the deterministic correlation function and $\hat{r}_v(m)$ the normalized correlation function estimate

$$r_h(m) = \frac{\sum_{l=-\infty}^{\infty} h(l)h(l+m)}{\sum_{l=-\infty}^{\infty} h(l)^2} \quad , \quad \hat{r}_v(m) = \frac{\sum_{n=1}^{N} v(n)v(n+m)}{\sum_{n=1}^{N} v(n)^2}$$

In [53] it is shown that

$$\lim_{N\to\infty} \hat{r}_v(m) = r_h(m) \tag{3.13}$$

This surprising result is further generalized to the fourth order moment and cumulant functions in [86]. The consequences of this are that many of the algorithms for regression and parameter estimation can be modified to be applicable to linear αS processes.

3.3.2 Covariation

The *covariance* function is a very powerful tool in the study of Gaussian random variables, but is not defined for non-Gaussian αS variables. For SαS random variables with $1 < \alpha < 2$ the *covariation* plays an analogous role to the covariance. Unfortunately, it is not as powerful, because it lacks some of the desirable properties of the covariance. It is, however, a useful quantity. Before defining the covariation, we introduce the signed power. The signed power p for a number u is defined as

$$u^{\langle p\rangle} = \begin{cases} |u|^p \mathrm{sign}(u) & \text{for } u \in \mathbb{R} \\ |u|^{p-1}u^* & \text{for } u \in \mathbb{C} \end{cases} \tag{3.14}$$

Now we define the covariation and the covariation coefficient

DEFINITION 3.4 *For jointly SαS random variables u_1 and u_2 with $1 < \alpha \leq 2$, the covariation of u_1 with u_2 is*

$$[u_1, u_2]_\alpha = \int_S u_1 u_2^{\langle\alpha-1\rangle} d\Gamma(\mathbf{s})$$

where $\Gamma(\mathbf{s})$ is the spectral measure of the SαS random vector $[u_1\ u_2]^T$.

DEFINITION 3.5 *For jointly SαS random variables u and v, the covariation coefficient of u with v is*

$$\lambda_{uv} = \frac{[u, v]_\alpha}{[v, v]_\alpha}$$

[67, 73]. Notice that the covariation coefficient is not normalized to the interval $[-1\ ;\ 1]$. The definition of the covariation $[u_1, u_2]_\alpha$ involves the spectral measure which appears in the spectral representation of the characteristic function of the vector $[u_1\ u_2]^T$. Because of that the definition of the covariation is not

very convenient in practice. Fortunately the covariation and the covariation coefficient are related to the fractional lower order moment. Let u and v be jointly SαS with $1 < \alpha \leq 2$, then

$$\begin{aligned}
[v, v]_\alpha &= \gamma_v^\alpha \\
\lambda_{uv} &= \frac{\mathrm{E}\left(uv^{\langle p-1 \rangle}\right)}{\mathrm{E}\left(|v|^p\right)} \quad , \quad 1 \leq p < \alpha \\
[u, v]_\alpha &= \frac{\mathrm{E}\left(uv^{\langle p-1 \rangle}\right)}{\mathrm{E}\left(|v|^p\right)} \gamma_v^\alpha \quad , \quad 1 \leq p < \alpha
\end{aligned} \tag{3.15}$$

[20, 18]. From this a practical estimator of the covariation based on the fractional lower order moment is readily obtained by replacing the expectation values with the empirical mean values.

The covariation function is linear in the first argument, this is sometimes referred to as *left linear.*

PROPERTY 3.6 *For SαS variables the covariation $[u, v]_\alpha$ is linear with respect to u. If u is a linear combination of SαS random variables, $u = a_1 u_1 + a_2 u_2$, where u_1,u_2, and v are jointly SαS , then*

$$[u, v]_\alpha = a_1 [u_1, v]_\alpha + a_2 [u_2, v]_\alpha$$

The covariation function is non-linear in the second argument, this is sometimes referred to as *right pseudo linear*

PROPERTY 3.7 *For SαS variables the covariation $[u, v]_\alpha$ possesses a certain linearity property with respect to v. If v is a linear combination of independent SαS random variables, $v = a_1 v_1 + a_2 u_2$, where v_1, v_2 are independent and u,v_1, and v_2 are jointly SαS , then*

$$[u, v]_\alpha = a_1^{\langle \alpha-1 \rangle} [u, v_1]_\alpha + a_2^{\langle \alpha-1 \rangle} [u, v_2]_\alpha$$

[20]. If u and v are independent, then $[u, v]_\alpha = 0$, while the converse is not true in general. For independent SαS variables the covariation separates [84]

PROPERTY 3.8 *Let u, v_1, v_2 be jointly SαS variables, with v_1 and v_2 independent, then the covariation $[u, v_1 + v_2]_\alpha$ is*

$$[u, v_1 + v_2]_\alpha = [u, v_1]_\alpha + [u, v_2]_\alpha$$

Example: Covariation for Linear SαS Process Consider a SαS linear process (MA process) with coefficients a_i, independent input sequence v_n following a SαS distribution with unit dispersion $\gamma_u = 1$, and an output sequence v_n

$$v_n = \sum_{l=0}^{L} a_l u_{n-l}$$

It is readily shown using Property 3.6 and 3.7 that the covariation is related to the coefficients by

$$[v_n, v_{n+k}]_\alpha = \sum_{l=0}^{L-k} \frac{a_l}{a_{k+l}} |a_{k+l}|^\alpha \qquad \text{and} \qquad [v_n, v_{n-k}]_\alpha = \sum_{l=0}^{L-k} \frac{a_{k+l}}{a_l} |a_l|^\alpha$$

for $k = 1, \ldots, L$. The dispersion of the process is $\gamma_v = [v_n, v_n]_\alpha = \sum_{l=0}^{L-k} |a_l|^\alpha$.

3.3.3 Estimation of Distribution Parameters

This section considers two specific methods for estimation of distribution parameters, one for estimation of parameters in the SαS distribution and one for the NIG distribution. Several estimators with different properties are found in the literature, see [66, 89, 82, 46] and references herein.

The estimators considered here are on a closed form and are relatively simple to calculate, which is often desirable in the context of signal processing. The cost is that the estimators require relatively large data sets, typically > 1000 samples, to obtain good estimates.

Estimation of Parameters in the SαS Distribution In [66] a parameter estimator, based on the so-called $\ln|\mathrm{S}\alpha\mathrm{S}|$-process, for the S$\alpha$S distribution is proposed. To introduce the concept of the $\ln|\mathrm{S}\alpha\mathrm{S}|$-process and because the estimator is easy to derive, the complete derivation will be given here.

Recall that the absolute pth moment of a SαS random variable satisfies Equation 3.12 for $-1 < p < \alpha$. Let x be a random SαS(α, γ) variable, and define a new variable $y = \ln|x|$. The absolute pth order moment of x is

$$\mathrm{E}\left(|x|^p\right) = \mathrm{E}\left(\exp(py)\right) \quad , \quad -1 < p < \alpha$$

and this is simply the moment generating function for the random variable y. Expanding the exponential function into power series gives

$$\mathrm{E}\left(\exp(py)\right) = \sum_{k=0}^{\infty} \mathrm{E}\left(y^k\right) \frac{p^k}{k!}$$

The higher order moments of y exist and they satisfy

$$\mathrm{E}\left(y^k\right) = \frac{d^k}{dp^k} \left(C_1(p,\alpha)\gamma^{p/\alpha}\right)\Big|_{p=0}$$

[66].Simplifying the above equation gives

$$\text{(a)} \quad \mu_y = C_e\left(\frac{1}{\alpha} - 1\right) + \frac{1}{\alpha}\ln\gamma \qquad\qquad \text{(b)} \quad \sigma_y^2 = \frac{\pi^2}{6}\left(\frac{1}{\alpha^2} + \frac{1}{2}\right) \tag{3.16}$$

where $\mu_y = \mathrm{E}\left(y\right)$, $\sigma_y^2 = \mathrm{E}\left((y - \mathrm{E}\left(y\right))^2\right)$, and $C_e = 0.57721566\ldots$ is the Euler constant. Based on mean and variance estimates of y, and rearranging Equation 3.16(a) and Equation 3.16(b) the parameter estimator is

$$\text{(a)} \quad \hat{\alpha} = \left(\frac{6}{\pi^2}\hat{\sigma}_y^2 - \frac{1}{2}\right)^{-1/2} \qquad\qquad \text{(b)} \quad \hat{\gamma} = \exp\left(\left(\hat{\mu}_y - \left(\frac{1}{\hat{\alpha}} - 1\right)C_e\right)\hat{\alpha}\right) \tag{3.17}$$

where $\hat{\sigma}_y^2$ and $\hat{\mu}_y$ are the estimate of the mean and variance of y, and $\hat{\alpha}$ in Equation 3.17(b) is the estimate of the characteristic exponent from Equation 3.17(a).

Estimation of Parameters in the Normal Inverse Gaussian Distribution In [46] a cumulant based parameter estimator of the parameters in the NIG distribution is proposed. The estimator takes advantage of the simple form of the chf.. The complete derivation of the cumulant based estimator is given, because it give some useful insight into the nature of the NIG distribution.

From the chf. in Equation 3.8 the cgf. of a random NIG$(\alpha, \beta, \delta, \mu)$ distributed variable is

$$\psi_v(t) = \delta\sqrt{\alpha^2 - \beta^2} - \delta\sqrt{\alpha^2 - (\beta + jt)^2} + j\mu t \tag{3.18}$$

The cumulants are readily obtained as the coefficients in the Taylor expansion of the cgf.. The first four cumulants are

$$\begin{aligned}
&\text{(a)} \quad c_1 = \mu + \frac{\beta\delta}{(\alpha^2-\beta^2)^{1/2}} && \text{(b)} \quad c_2 = \frac{\alpha^2\delta}{(\alpha^2-\beta^2)^{3/2}} \\
&\text{(d)} \quad c_3 = \frac{3\alpha^2\beta\delta}{(\alpha^2-\beta^2)^{5/2}} && \text{(d)} \quad c_4 = \frac{3\alpha^2(\alpha^2+4\beta^2)\delta}{(\alpha^2-\beta^2)^{7/2}}
\end{aligned} \tag{3.19}$$

Notice that contrary to the αS distribution the cumulants exist and are algebraic expressions of the parameters. Furthermore the skewness and kurtosis have the following closed form expressions

$$\text{(a)} \quad \gamma_3 = \frac{c_3}{(c_2)^{3/2}} = \frac{3\rho}{\xi^{1/2}} \qquad\qquad \text{(b)} \quad \gamma_4 = \frac{c_4}{c_2^2} = \frac{3(1+4\rho^2)}{\xi} \tag{3.20}$$

where $\rho = \beta/\alpha$ and $\xi = \delta(\alpha^2-\beta^2)^{1/2}$. Notice that, due to the permissible range of α and β, $-1 < \rho < 1$ and $0 < \xi < \infty$. Combining the expression of the skewness and kurtosis it is readily shown that the NIG distribution can model distributions where $\gamma_4 \geq 4\gamma_3^2/3$.

The four cumulants in Equation 3.19(a)-(d) depend explicitly on the four parameters in the NIG distribution. The cumulant based parameter estimator is obtained by a combination of the cumulants in Equation 3.19 and the expressions for the skewness and kurtosis in Equation 3.20. From Equation 3.20 we obtain

$$\text{(a)} \quad \xi = \left(\gamma_4 - \frac{4}{3}\gamma_3^2\right)^{-1} \qquad\qquad \text{(b)} \quad \rho = \frac{\gamma_3}{3}\xi^{1/2} \tag{3.21}$$

Combining Equation 3.19 and Equation 3.21 the NIG parameters have the following algebraic expressions

$$\begin{aligned}
&\text{(a)} \quad \alpha = \frac{\xi}{\delta(1-\rho^2)^{1/2}} && \text{(b)} \quad \beta = \alpha\rho \\
&\text{(d)} \quad \delta = \left(c_2\xi(1-\rho^2)\right) && \text{(d)} \quad \mu = \rho(c_2\xi)^{1/2}
\end{aligned} \tag{3.22}$$

The parameter estimator can now be constructed from the above equations as follows: (1) Estimate the four cumulants (2) Based on the estimated cumulants calculate the skewness and kurtosis using Equation 3.20 (3) Insert the estimated cumulants and calculated values in Equation 3.22(a), 3.22(b), 3.22(c) and 3.22(d) in precisely that order.

3.3.4 Estimation of the Spectral Measure

From Theorem 3.2 we know that the joint distribution of an αS vector is characterized by the spectral measure and the location vector. This section considers a method for the estimation of the spectral measure and the performance of the estimation procedure subject to the sampling radius and the characteristic exponent is discussed.

Discrete Approximation of the Spectral Measure From [17, Theorem 1] we know that for an M-dimensional αS variable with spectral measure $\Gamma(\mathbf{s})$ and density function $p(\mathbf{x})$, there is a discrete spectral measure $\Gamma_a(\mathbf{s})$ with corresponding density function $p_a(\mathbf{x})$ satisfying

$$\sup_{\mathbf{x}\in\mathbb{R}^M} |p(\mathbf{x}) - p_a(\mathbf{x})| \leq \epsilon$$

where $\epsilon > 0$. Thus an arbitrary spectral measure $\Gamma(\mathbf{s})$ can be approximated by a discrete measure $\Gamma_a(\mathbf{s})$ such that the corresponding densities are arbitrarily close. The only requirement is that sampling of $\mathbf{s}$ is sufficiently dense.

The chf. corresponding to the approximated discrete spectral measure $\Gamma_a(\mathbf{s})$ sampled in L points, can be written as

$$\Phi_a(\mathbf{t}) = \exp\left(-\sum_{n=1}^{L} \psi_\alpha(\mathbf{t}^T\mathbf{s}_n)\Gamma_a(\mathbf{s}_n)\right)$$

Now define a vector $\mathbf{\Gamma} = [\Gamma_a(\mathbf{s}_1)\ldots\Gamma_a(\mathbf{s}_L)]^T$ containing the L values of the approximated discrete spectral measure. Evaluating the approximated chf. for L values of $\mathbf{t}$ and defining the function $I_a(\mathbf{t}) = -\ln\Phi_a(\mathbf{t})$, we can formulate the set of linear equations

$$\mathbf{I} = \Psi\mathbf{\Gamma} \tag{3.23}$$

where

$$\mathbf{I} = \begin{bmatrix} I_a(\mathbf{t}_1) \\ \vdots \\ I_a(\mathbf{t}_L) \end{bmatrix} \quad , \quad \Psi = \begin{bmatrix} \psi_\alpha(\mathbf{t}_1^T\mathbf{s}_1) & \ldots & \psi_\alpha(\mathbf{t}_1^T\mathbf{s}_L) \\ \vdots & \ddots & \vdots \\ \psi_\alpha(\mathbf{t}_L^T\mathbf{s}_1) & \ldots & \psi_\alpha(\mathbf{t}_L^T\mathbf{s}_L) \end{bmatrix}$$

If the L values of $\mathbf{t}$ are chosen so that Ψ^{-1} exists, then the approximated spectral measure is given exactly by the solution to Equation 3.23.

Estimation of the Discrete Spectral Measure In [74] a method for estimating the spectral measure is proposed. The principle behind the estimation is based on Equation 3.23. From this point, without loss of generality and to simplify the presentation, we will assume that $\mathbf{x}$ is SαS . For estimation of the spectral measure in the non-symmetric case see the more general treatment in [74].

In the case of a symmetric pdf. the corresponding spectral measure and chf. is real valued and symmetric. From the definition of the chf. an estimate based on samples of the random vector $\mathbf{x}$ can be obtained as

$$\hat{\Phi}(\mathbf{t}) = \frac{1}{K}\sum_{k=1}^{K} \cos(\mathbf{t}^T\mathbf{x}_k) \tag{3.24}$$

for K samples of $\mathbf{x}$. By introducing the function $\hat{I}(\mathbf{t}) = -\ln(\hat{\Phi}_\alpha(\mathbf{t}))$ and substitute $\mathbf{I}$ by $\hat{\mathbf{I}} = [\hat{I}(\mathbf{t}_1)\ldots \hat{I}(\mathbf{t}_L)]^T$, an estimate of the approximate discrete spectral measure is directly obtained from Equation 3.23.

The spectral measure is defined on the M-dimensional unit sphere. If no a priori knowledge about the spectral measure is available and if all directions are of equal importance, it is natural to sample $\mathbf{s}$ as uniformly on the unit sphere as possible. In the SαS case it suffices to sample on the half M-dimensional unit sphere. For $M = 2$ the sampling is on the unit circle and uniform or equidistant sampling is readily obtained with $2\pi/L$ radians distance. For $M = 2$ and SαS the nth sample point in $\mathbf{s}$ is then $\mathbf{s}_n = [\cos\left(\pi\frac{n-1}{L}\right) \quad , \quad \sin\left(\pi\frac{n-1}{L}\right)]^T$. However, uniform sampling is in general not possible in dimensions higher than 1. For $M = 3$ the sampling is on the surface of the 3-dimensional unit sphere. Uniform sampling on the surface of a 3-dimensional sphere only exist for 5 specific number of sampling points[3]. Approximately uniform sampling in higher dimension is not trivial.

[3]Uniform sampling on the 3-dimensional unit sphere has the following solutions: 4 points $\sim$ tetrahedron, 6 points $\sim$ octahedron, 8 points $\sim$ cube, 12 points $\sim$ isocahedron, and 20 points $\sim$ dodecahedron

The natural choice of the sampling grid of the chf. is to sample symmetrically on a M-dimensional sphere, again in the SαS case it suffices to sample on a half M-dimensional sphere. Thus the nth sample point in $\mathbf{t}$ is $\mathbf{t}_n = r\ \mathbf{s}_n$, where r is the sampling radius. The sampling radius is decisive for the performance of the estimator of the spectral measure, this is discussed later in this section.

To avoid negative values in the estimate of the spectral measure $\mathbf{\Gamma}$, the solution in Equation 3.23 is restated as

$$\min \left\| \hat{\mathbf{I}} - 2\mathrm{Re}(\mathbf{\Psi})\mathbf{\Gamma} \right\|_2 \quad \text{s.t.} \ \mathbf{\Gamma} \geq \mathbf{0} \tag{3.25}$$

The estimation procedure for the spectral measure is now: (1) determine the sampling radius of $\mathbf{t}$ (2) calculate $\mathbf{\Psi}$ (3) estimate $\hat{\mathbf{I}}$ and (4) solve the constrained least square problem in Equation 3.25.

Performance of the Estimation Procedure Consider a random SαS vector $\mathbf{x} \in \mathbb{R}^2$ with the discrete spectral measure given in the table in Figure 3.3(a). Sampling uniformly on the half unit circle, and with $L = 90$ sampling points, gives the discrete spectral measure

$$\Gamma(\mathbf{s}_n) = \begin{cases} \frac{\gamma}{2} & \text{for } n = 1, 46\ (91, 136) \\ 0 & \text{else} \end{cases} \tag{3.26}$$

where $\gamma = 1$. Applying the estimation procedure we can estimate the spectral measure for the SαS random vector. In Figure 3.6 a typical example of an estimated spectral measure is shown. Observe

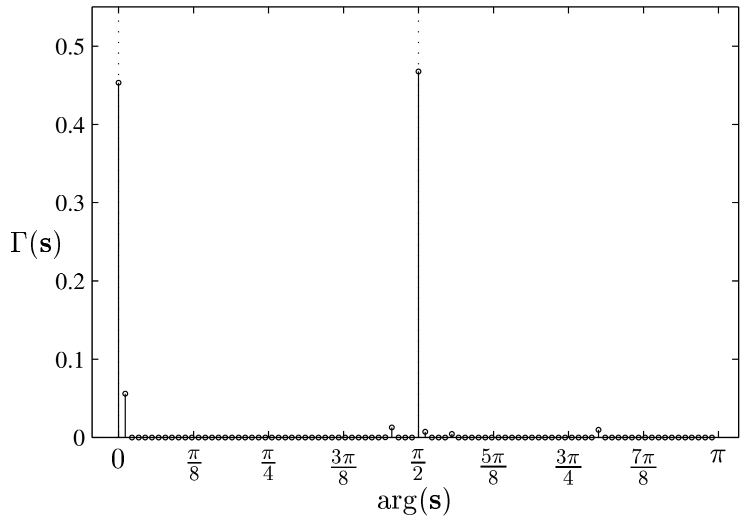


Figure 3.6 Estimated spectral measure, $\Gamma(\mathbf{s})$, for SαS$(1, 1)$ random vector. The estimation is based on $K = 10000$ observations of the random vector $\mathbf{x}$, sampling radius $r = \frac{1}{2}$, and $L = 90$. The direction in which the true spectral measure has amplitude are indicated by dotted lines.

that the distribution of masses for the estimated spectral measure is very specific in the 2 directions corresponding to the directions in which the true spectral measure has masses.

It is interesting to examine the statistical characteristics of the estimator subject to the sampling radius of the chf., and the choice of the characteristic exponent. To evaluate the performance of the estimation of the spectral measure we define the error function

$$\Upsilon = \mathrm{E}\left((\mathbf{I} - \hat{\mathbf{I}})^T (\mathbf{I} - \hat{\mathbf{I}}) \right) / (\mathbf{I}^T \mathbf{I}) \tag{3.27}$$

where $\mathbf{I}$ is defined in Equation 3.23 and $\hat{\mathbf{I}}$ is the estimate of $\mathbf{I}$ that corresponds to the estimate of the spectral measure from Equation 3.25.

The sampling radius of the chf. has crucial influence on the estimator of the spectral measure. A desirable property is to have an estimator that is independent of the scaling of $\mathbf{x}$. From Equation 3.3 and Equation 3.5 one directly obtains that if $\gamma = \gamma_n$, $n = 1, \ldots, N$ and

$$r \propto (N\gamma)^{-1/\alpha} \tag{3.28}$$

then the estimator is scale independent[4].

Consider again the SαS random vector with the spectral measure in Equation 3.26. In Figure 3.7(a) the error function versus the sampling radius is shown for $\gamma = 0.01, 0.1, 1, 10$. The vertical bars indicate

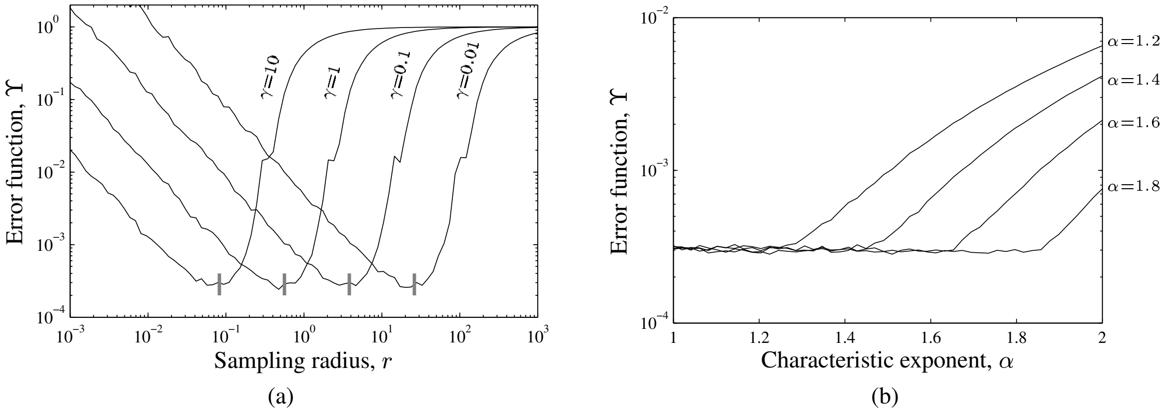


Figure 3.7 Error function for the estimation of the spectral measure. (a) Error function versus sampling radius, $\mathbf{x} \in \mathrm{S}\alpha\mathrm{S}(1.2, \gamma)$. (b) Error function versus the choice of characteristic exponent in the estimation. $\mathbf{x} \in \mathrm{S}\alpha\mathrm{S}(\alpha, 1)$.

the sampling radius corresponding to $r = (N\gamma)^{-1/\alpha}$, and it appears that this is an appropriate sampling radius. In Figure 3.7(b) the error function versus the choice of the characteristic exponent is shown for $\mathbf{x} \in \mathrm{S}\alpha\mathrm{S}(\alpha, 1)$, and $\alpha = 1.2, 1.4, 1.6, 1.8$. The influence of the characteristic exponent is not as crucial as one could expect, and it appears that in cases where the characteristic exponent is unknown, it must be selected reasonably low. See also the examples in [56] included in Appendix A.4.

3.4 Summary

In this chapter several heavy tail distributions are introduced in a signal processing context, and special emphasis is on the class of αS distributions and the class of NIG distributions. Heavy tail stochastic processes are introduced and the relation between a time-domain and a frequency-domain description is discussed.

The temporal dependence structure of infinite variance stochastic processes cannot directly be revealed by second order moment functions and therefore the fractional lower order moment function, the normalized moments and cumulants, and the covariation function are introduced. Methods for estimation of parameters in the SαS and NIG distribution suitable in signal processing are reviewed. The multivariate αS distributions are characterized by a spectral measure; a method for estimating the spectral measure is introduced and the characteristics of the estimator subject to the characteristic exponent of the distribution and the sampling radius of the characteristic function are examined.

[4]In [74] the sampling radius is fixed to $r = 1$, and data is scaled with the median of the data.

The main contribution of this chapter is that the theory and estimation methods for heavy tail distributions and processes are reviewed in a signal processing context.

CHAPTER 4

BLIND SIGNAL SEPARATION

The fundamentals of blind signal separation have been established in the literature through the last decade. This chapter reviews some of the most important methods for blind signal separation and special attention is paid to the implications of source signals with heavy tail distributions. Several new methods are proposed and simulation examples exhibit the properties of the methods.

The chapter is organized as follows. In Section 4.1 blind signal separation methods based on the maximum likelihood estimation principle are considered, and the maximum likelihood estimators are derived for instantaneous and convolutive mixtures. The blind signal separation problem has several formulations as joint diagonalization of matrices that exploits the instantaneous or temporal probability structure of the observable signals. In Section 4.2 some of the methods based on joint diagonalization of matrices are considered, and a unifying method that combines the existing methods is proposed. Finally, in Section 4.3 a method for blind signal separation based on the spectral measure for αS signals is proposed. The generative model is an instantaneous mixture and the method estimates the bases in the mixture without a priori knowledge on the dimension of the mixture.

4.1 Maximum Likelihood Methods

This section considers methods for blind signal separation based on a probabilistic estimation principle - the maximum likelihood method. Recently the blind signal separation problem has received considerable attention and the following introduction serves to relate this work to the vast of work in this field.

4.1.1 Introduction

Statistical principles for blind signal separation have evolved from the field of Independent Component Analysis (ICA) with Jutten et. al 1991 [51, 28] among others as early pioneering works. It is appropriate to mention a few more key-references on the trajectory to the present state of the art. Comon 1994 [27] derived an ICA algorithm based on the 4.th order Edgeworth expansion of the negentropy. Bell and Sejnowski 1995 [10] derived a stochastic gradient ICA method based on an information maximization approach. Cardoso 1998 [22] reviewed the ICA methods from a maximum likelihood principles point of view, and Attias and Schreiner 1998 [6] extended the ICA to exploit the temporal dependence structure of the source signals. Recently Lee et. al. [60] reviewed the field of ICA and established a unifying information-theoretic framework. Even though [60] does not include the methods in [6], the thorough treatment makes [60] a good and recommendable survey of ICA methods.

To establish a probabilistic estimation model for the blind signal separation problem it is useful to formulate a generative model of the observable signals. The observable signals are arranged into an observation vector $\mathbf{x}$ characterized by the pdf. $f_{\mathbf{x}}(\mathbf{x})$. The observable vector is assumed to be generated from a latent source vector of independent signals, and the mixture model describes the transformation from the source vector to the observable vector. From the generative model a pdf. model $f_m(\mathbf{x})$ of the observation vector can be formulated.

A suitable measure of the distance between the pdf. of the observable vector $f_{\mathbf{x}}(\mathbf{x})$ and the pdf. model of the the observable vector $f_m(\mathbf{x})$ is the Kullback-Leibler distance

$$\begin{aligned} D\left(f_{\mathbf{x}}(\mathbf{x})\,\|f_m(\mathbf{x})\right) &= \int f_{\mathbf{x}}(\mathbf{x}) \ln \frac{f_{\mathbf{x}}(\mathbf{x})}{f_m(\mathbf{x})} d\mathbf{x} \\ &= \underbrace{\int f_{\mathbf{x}}(\mathbf{x}) \ln f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}}_{\text{Entropy}} - \underbrace{\int f_{\mathbf{x}}(\mathbf{x}) \ln f_m(\mathbf{x}) d\mathbf{x}}_{\text{log-likelihood}} \end{aligned} \tag{4.1}$$

The first term on the right hand side is the entropy of the observations and the second term is the log-likelihood of the parametric pdf. model. Thus minimizing the distance between the observations and the pdf. model is to maximize the log-likelihood of the parametric pdf. model.

In the following the maximum likelihood estimation is derived for three different mixture models, with special emphasis on sources with heavy tail distributions. In Section 4.1.2 the maximum likelihood estimation of an instantaneous mixture of temporal independent sources is derived, and in Section 4.1.3 the estimation is extended to an instantaneous mixture of temporal dependent sources. Finally in Section 4.1.4 the maximum likelihood estimation for a convolutive mixture of temporal independent sources is derived.

4.1.2 Instantaneous Mixtures of Temporal Independent Sources

In [10] Bell and Sejnowski proposed the information-maximization ICA method and derived a stochastic gradient estimation method based on a logistic non-linear function. Cardoso [21] formulated the ICA problem as a maximum likelihood problem and showed that the information-maximization and the maximum likelihood formulations are identical. The simplest generative model is an instantaneous mixture of temporal independent sources, and this was the generative model of the early works [10, 21]. In the following the stochastic gradient maximum likelihood ICA method is derived for the instantaneous mixture, and a discussion of the effects of model deviations and heavy tail source distributions is provided.

Consider the mixture model $\mathbf{x} = \mathbf{A}\mathbf{s}$ where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the mixture matrix, $\mathbf{x} = [x_1, \ldots, x_N]^T$ is the observable vector and $\mathbf{s} = [s_1, \ldots, s_N]^T$ is the source vector. In addition to the generative model it is useful to define an estimation model $\mathbf{u} = \mathbf{D}\mathbf{x}$, where $\mathbf{D}$ is an estimate of the inverse of the mixture matrix and $\mathbf{u} = [u_1, \ldots, u_N]^T$ is the estimate of the source vector. The mixture and estimation models are depicted in Figure 4.1.

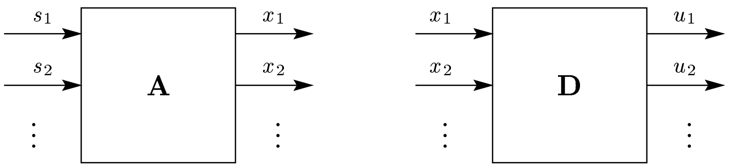


Figure 4.1 Left: The mixture model $\mathbf{x} = \mathbf{A}\mathbf{s}$. Right: The estimation model $\mathbf{u} = \mathbf{D}\mathbf{x}$.

It appears that if $\mathbf{D} = \mathbf{A}^{-1}$ then the pdf. of the source vector $f_{\mathbf{s}}(\mathbf{s})$ is equal to the pdf. of the output vector $f_{\mathbf{u}}(\mathbf{u})$, and since the elements in the source vector are independent, the joint pdf. of the output vector is the product over the marginal pdf.s

$$f_{\mathbf{u}}(\mathbf{u}) = \prod_{n=1}^{N} f_{u_n}(u_n)$$

where $f_{u_n}(u_n)$ is the marginal pdf. of the nth element of $\mathbf{u}$. The relation between the pdf. of a random vector and the pdf. of a linear transformation of the random vectors is given in Equation 2.13. Now the estimation model of the observable vector $\mathbf{x}$ in terms of $\mathbf{D}$ can be written as

$$f_m(\mathbf{x}, \mathbf{D}) = |\det(\mathbf{D})| \prod_{n=1}^{N} f_{u_n}(u_n) \tag{4.2}$$

If the source distribution is known a priori this equation is the pdf. of the observed vector parameterized with $\mathbf{D}$. The log-likelihood function in Equation 4.1 is the expectation, over observations of $\mathbf{x}$, of the pdf. model $f_m(\mathbf{x}, \mathbf{D})$. Thus, for the instantaneous mixture of temporal independent sources, the likelihood of $\mathbf{D}$ is

$$\mathcal{L}(\mathbf{D}) = \mathrm{E}\left(\ln|\det(\mathbf{D})| + \sum_{n=1}^{N} \ln f_{u_n}(u_n)\right) \tag{4.3}$$

The maximum likelihood estimator of $\mathbf{D}$ is the maximum of Equation 4.3. The first step towards a development of an estimator is to take the derivative of the log-likelihood function in Equation 4.3 with respect to $\mathbf{D}$. The derivative of the first term in Equation 4.3 is

$$\frac{\partial}{\partial \mathbf{D}} \ln|\det(\mathbf{D})| = \frac{(\mathrm{adj}\,\mathbf{D})^T}{\det \mathbf{D}} = \left(\mathbf{D}^T\right)^{-1}$$

The derivative of the second term in Equation 4.3 with respect to the (m, n)th element of $\mathbf{D}$ is

$$\frac{\partial}{\partial \mathbf{D}_{mn}} \sum_{n=1}^{N} \ln f_{u_n}(u_n) = \frac{\partial}{\partial u_m} \ln f_{u_m}(u_m) \frac{\partial u_m}{\partial \mathbf{D}_{mn}} = \frac{1}{f_{u_m}(u_m)} \frac{\partial f_{u_m}(u_m)}{\partial u_m} x_n$$

Now the derivative of the log-likelihood function with respect to $\mathbf{D}$ can be written as

$$\frac{\partial}{\partial \mathbf{D}} \mathcal{L}(\mathbf{D}) = \left(\mathbf{D}^T\right)^{-1} + \mathrm{E}\left(g(\mathbf{u})\mathbf{x}^T\right) \tag{4.4}$$

where $g(\mathbf{u})$ is a non-linear function operating element-wise on the output vector, i.e. $g(\mathbf{u}) = [g(u_1) \ldots g(u_N)]^T$. The non-linear function g appears from the second term of the derivative of the log-likelihood function, and is given by

$$g(x) = \frac{1}{f_m(x)} \frac{\partial f_m(x)}{\partial x} \tag{4.5}$$

where $f_m(x)$ is the pdf. model of the source signal.

Examples of Likelihood Functions The likelihood function in Equation 4.3 is parameterized by the pdf. model of the sources, and to maximize the likelihood function a priori assumptions on the pdf. of the sources are necessary. The likelihood function must exhibit robustness against deviations between model and source distribution, and the following graphical interpretation of the likelihood function illustrates the influence of deviations between the model and the source distribution[1].

[1]This example is inspired by a similar example in [22].

Let the model distribution of the sources be the double exponential pdf.. For simplicity assume that the mixture model is $\mathbf{A} = \mathbf{I}$ and consequently the likelihood function is expected to have its maximum for $\mathbf{D} = \mathbf{I}$. To generate a neighborhood of the identity matrix the following parameterization of the estimation matrix is convenient

$$\mathbf{D}(u,v) = \begin{bmatrix} \cosh(u) & \sinh(u) \\ \sinh(u) & \cosh(u) \end{bmatrix} \begin{bmatrix} \cos(v) & -\sin(v) \\ \sin(v) & \cos(v) \end{bmatrix} \tag{4.6}$$

Observe that $\mathbf{D}(0,0) = \mathbf{I}$. The parameter u controls an anti-diagonal symmetric deviation and v controls an anti-diagonal skew-symmetric deviation of $\mathbf{D}(u,v)$ away from the identity matrix.

The likelihood function is depicted for 4 different source distributions in Figure 4.2. In Figure 4.2(a) the source distribution is identical with the model distribution and as expected the likelihood function has maximum for $\mathbf{D}(0,0) = \mathbf{I}$. In Figure 4.2(b) and Figure 4.2(c) the source distribution is SαS and symmetric NIG respectively, and again it appears that the likelihood function has maximum for $\mathbf{D} = \mathbf{I}$. Thus, even though there is a considerable deviation between the model distribution and the source distribution the likelihood function has a maximum at the right estimation matrix. The general experience

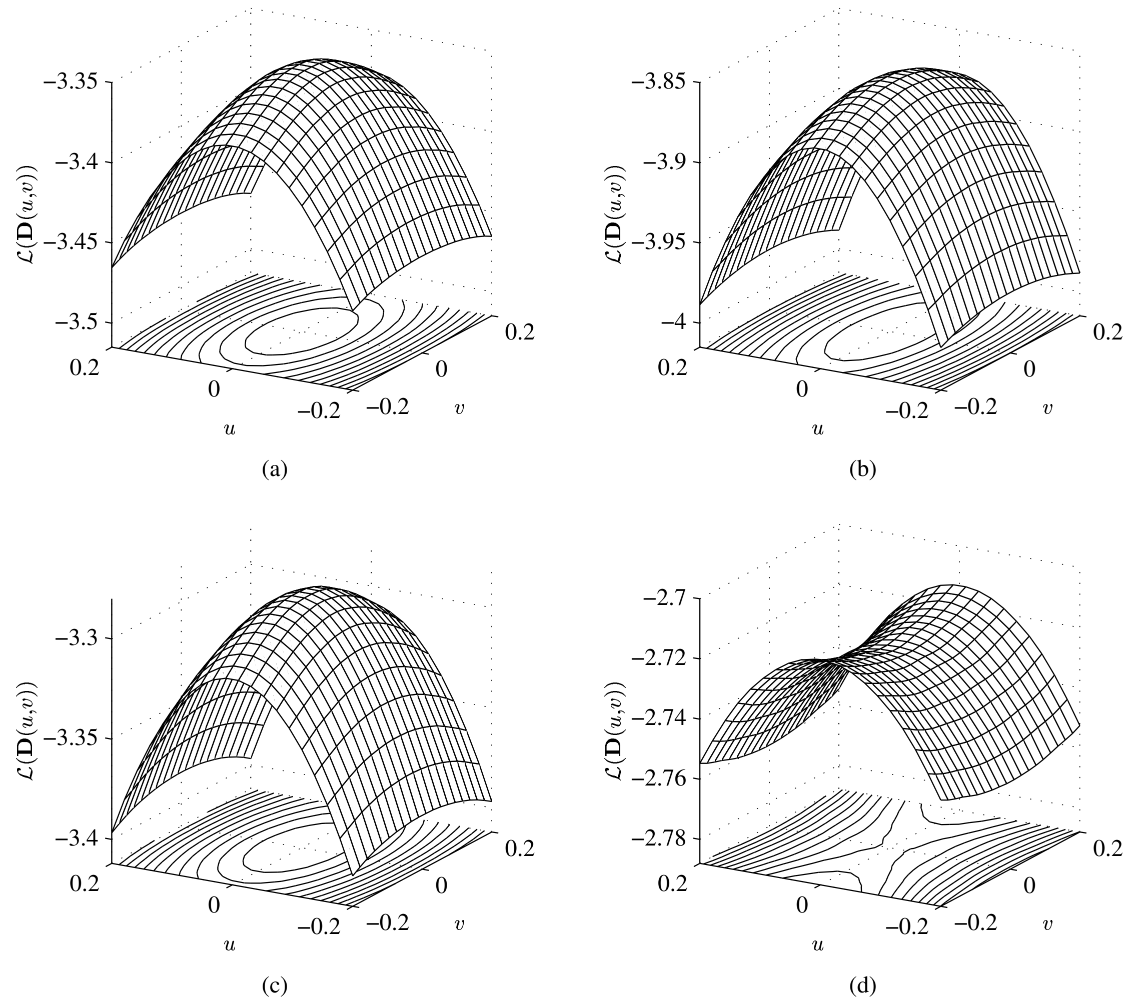


Figure 4.2 Plots of log-likelihood functions, the model distribution is the double exponential pdf. and the source distributions are: (a) double exponential, (b) SαS$(1.6, 1)$, (c) NIG$(0.5, 0, 1, 0)$, and (d) uniform.

is that as long as the model distribution is reasonably close to the source distribution, the likelihood function has maximum for $\mathbf{D} = \mathbf{A}^{-1}$. In Figure 4.2(d) the likelihood function is depicted for a uniform source distribution, and it appears that the likelihood does not have maximum for $\mathbf{D} = \mathbf{I}$. This is an example where the deviation between the model and the source distribution is too large.

To the knowledge of the author this is the first examination of the likelihood functions for the blind signal separation problem for SαS and NIG distributed sources.

Adaptation Equations Using the gradient of the likelihood in Equation 4.4 it is straightforward to develop stochastic gradient algorithms for the estimation of $\mathbf{D}$. The stochastic gradient algorithms use an iterative updating of $\mathbf{D}$ on the form $\mathbf{D}_{k+1} = \mathbf{D}_k + \mu\Delta\mathbf{D}_k$, where $\Delta\mathbf{D}$ is some kind of gradient and μ is an adaptation constant. From the derivative of the likelihood function in Equation 4.4 the *ordinary* gradient is

$$\Delta\mathbf{D} = \left(\mathbf{D}^T\right)^{-1} + \frac{1}{K}\sum_{k=1}^{K} g(\mathbf{u})\mathbf{x}^T \tag{4.7}$$

where the expectation is replaced with an averaging over K samples. The iterative updating can be performed *instantaneously* corresponding to $K = 1$, or the gradient can be estimated over an ensemble of observations, $K > 1$.

While the ordinary gradient adaptation is efficient in Euclidean matrix spaces, it may have poor convergence properties in non-Euclidean spaces. In general $\mathbf{D}$ forms a non-Euclidean matrix space, and algorithms with more efficient convergence properties can be obtained by exploring the structure of the parameter space. Another gradient, the *natural* gradient or the *relative* gradient, proposed independently in [5] and [23], has in general better convergence properties for the iterative updating of $\mathbf{D}$. The relative gradient is defined as

$$\text{relative gradient} = \frac{\partial\mathcal{L}(\mathbf{D})}{\partial\mathbf{D}}\mathbf{D}^T\mathbf{D}$$

Multiplying the ordinary gradient with $\mathbf{D}^T\mathbf{D}$ the relative gradient update is readily obtained

$$\Delta\mathbf{D} = \left(\mathbf{I} + \frac{1}{K}\sum_{k=1}^{K} g(\mathbf{u})\mathbf{u}^T\right)\mathbf{D} \tag{4.8}$$

Notice that the relative gradient update does not imply inversion of $\mathbf{D}$, which makes the relative gradient update computational more attractive than the ordinary gradient update[2].

Depiction of the Relative Gradient Figure 4.3 shows the contours of the likelihood function from Figure 4.2(a), which is the likelihood function on the manifold described by Equation 4.6. It is instructive to depict the relative gradient of the log-likelihood function, and the arrows in Figure 4.3 indicate the (approximate) relative gradient. Because the gradients in general will not be on the manifold described by Equation 4.6, it is only possible to depict approximate gradients on the manifold (minimum norm between the gradient and the approximate gradient on the manifold).

[2]For the development of efficient algorithms the work of [34, 33], where the geometry of algorithms with orthogonality constraints are investigated, should be noticed. These methods are recently applied to the blind signal separation problem in [32].

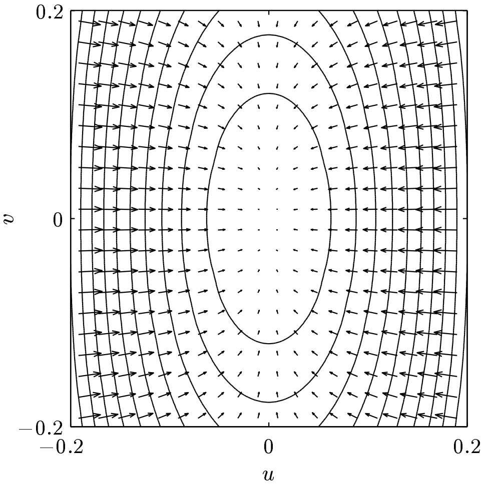


Figure 4.3 Contour of the log-likelihood function in Equation 4.3 for $\mathbf{D}$ parameterized as in Equation 4.6. The arrows indicate the (approximate) relative gradient.

Non-linear Functions for Heavy Tail Sources In the literature different non-linearities have been applied to the blind signal separation problem, see for instance [10, 42, 40, 48]. Among the most popular non-linearities are $g(x) = -\text{sign}(x)$ and $g(x) = -\tanh(x)$ corresponding to the double exponential pdf. $f(x) = \frac{1}{2}\exp(-|x|)$ and the inverse-cosine-hyperbolic pdf. $f(x) = \frac{1}{\pi}\frac{1}{\cosh(x)}$, respectively.

It is interesting to investigate the non-linear functions corresponding to the heavy tail pdf.s considered in Chapter 3. For the Gaussian and the Cauchy distributions Equation 4.5 yields $g(x) = -2x$ and $g(x) = -2x/(1+x^2)$ respectively. Even though closed forms of the SαS pdf.s are only known for the Gaussian and the Cauchy distributions, it is interesting to depict the non-linearities for other SαS pdf.s. In Figure 4.4(a) non-linearities for SαS pdf.s for 4 different values of the characteristic exponent are depicted.

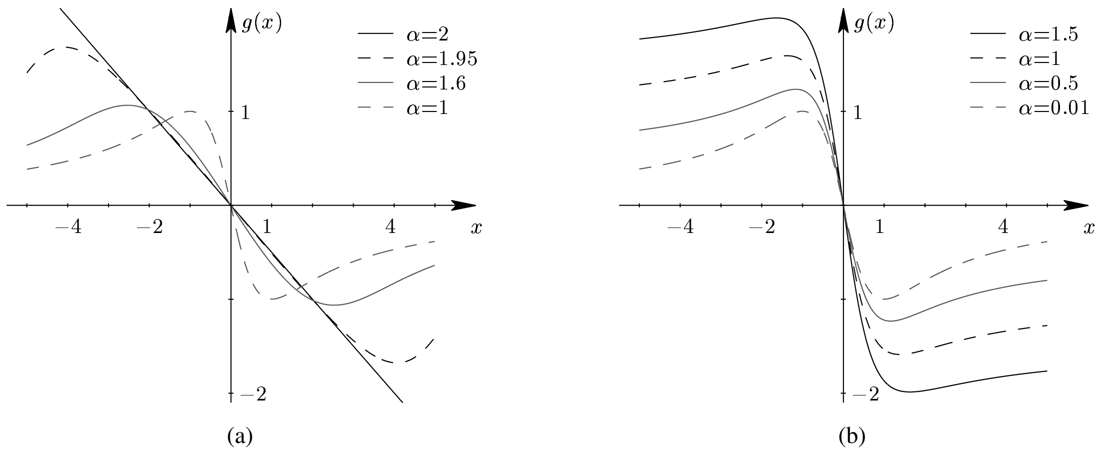


Figure 4.4 Non-linear function for heavy tail pdf.s. (a) Non-linear function for SαS pdf.s, $\gamma = 1$. (b) non-linear functions for NIG pdf.s, $\beta = 0$, $\delta = 1$.

For the class of symmetric NIG pdf.s it is straightforward to obtain[3]

$$g(x) = \frac{-\alpha x}{\sqrt{\delta^2 + x^2}} \frac{K_2(\alpha\sqrt{\delta^2 + x^2})}{K_1(\alpha\sqrt{\delta^2 + x^2})} \tag{4.9}$$

where $K_1(\cdot)$ and $K_2(\cdot)$ are the modified Bessel function of the second kind with index 1 and 2. In Figure 4.4(b) non-linearities for 4 different values of the tail parameter for NIG pdf.s are depicted. It is interesting to notice from Figure 4.4 that for all non-Gaussian NIG and SαS pdf.s the function $g(x)$ is a transform from an unbounded pdf. to a bounded pdf..

The second term of the gradient of the log-likelihood function in Equation 4.4 is the covariance between the observation signal and a non-linear function of the estimated signal. In the examination of the properties for the estimation of SαS sources, it is useful to consider the relative gradient

$$\frac{\partial}{\partial \mathbf{D}} \mathcal{L}(\mathbf{D})\mathbf{D}^T\mathbf{D} = \left(\mathbf{I} + \mathrm{E}\left(g(\mathbf{u})\mathbf{u}^T\right)\right)\mathbf{D} \tag{4.10}$$

Observe that if $g(u) = |u|^{p-1}\mathrm{sign}(u) = u^{\langle p-1\rangle}$ then the diagonal of $\mathrm{E}\left(g(\mathbf{u})\mathbf{u}^T\right)$ is the absolute pth order moment introduced in Equation 3.11, and $\mathrm{E}\left(g(\mathbf{u})\mathbf{u}^T\right)$ is a simple generalization to the absolute pth order spatial moment matrix. Thus for $g(u) = -|u|^{p-1}\mathrm{sign}(u)$ and $0 < p \leq \alpha - 1$ the gradient is well-defined even for αS sources, and for $p = 1$ the non-linear function is simply $g(u) = -\mathrm{sign}(u)$ which is the non-linearity obtained for a source with double exponential pdf..

From Figure 4.2 it appears that the likelihood function for the SαS and NIG pdf. has maximum for $\mathbf{D} = \mathbf{A}^{-1}$ if the simple sign-function is used as the non-linear function. Furthermore the gradient is well-defined even in the case of infinite variance source pdf.s, and thus the stochastic gradient method for estimation of $\mathbf{D}$ is a valid method for the estimation.

4.1.3 Instantaneous Mixtures of Temporal Dependent Sources

In [6] Attias and Schreiner proposed a method that extends the estimation model in Section 4.1.2 to exploit the temporal dependence structure of the source signals[4]. Most real world signals have temporal dependence and this a priori knowledge can by advantage be utilized in the estimation model. ICA methods such as the one described in Section 4.1.2 assume temporal independent sources and these methods in general exhibit slow convergence for sources with temporal dependency structure. By utilizing the temporal dependence structure of the sources, better convergence properties can be obtained and the condition of non-Gaussian sources can be relaxed. In the original work [6] three different methods are derived, a time-domain, a frequency domain, and a hybrid between the time- and frequency-domain method. The frequency-domain and the hybrid methods have better convergence properties, however in this work only the time-domain method is considered.

Let the generative model be defined as $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$, where k is a time-index, $\mathbf{x}_k = [x_{1,k}, \ldots, x_{N,k}]^T$ is the observable vector, $\mathbf{s}_k = [s_{1,k}, \ldots, s_{N,k}]^T$ is the source vector and $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the instantaneous mixture matrix. Moreover the source signals are modelled as stochastic processes. The nth source signal with time-index k is modelled as the AR-process

$$s_{n,k} = \sum_{l=1}^{L-1} a_{n,l}s_{n,k-l} + e_{n,k-l} \tag{4.11}$$

where $e_{n,k}$ denotes the nth excitation process and $a_{n,l}$ are the coefficients of the process. To complete the notation for the generative model, the excitation process is organized in the vector $\mathbf{e}_k = [e_{1,k}, \ldots, e_{N,k}]^T$ and the coefficients for the nth source signal are collected in the vector $\mathbf{a}_n = [a_{n,1}, \ldots, a_{n,L-1}]^T$.

[3] see [2] for the derivative of the modified Bessel function.

[4] The method is termed Dynamic Component Analysis, DCA.

As for the method described in Section 4.1.2 it becomes useful to define a corresponding estimation model $\mathbf{u}_k = \mathbf{D}\mathbf{x}_k$, where $\mathbf{u}_k = [u_{1,k}, \ldots, u_{N,k}]^T$ is the output vector with time-index k and $\mathbf{D} \in \mathbb{R}^{N \times N}$ is the inverse mixture matrix. The estimate of the nth excitation process is a linear filtering of the nth element of the output vector

$$v_{n,k} = \sum_{l=0}^{L-1} d_{n,l} u_{n,k-l} \tag{4.12}$$

which is the inverse model of the AR-process for the nth source signal in Equation 4.11. The filter coefficients of the output filter are collected in the vector $\mathbf{d}_n = [d_{n,0}, \ldots, d_{n,L-1}]^T$. For the later development of the maximum likelihood estimator the following matrix notation of the output filter is introduced

$$\begin{bmatrix} v_{n,k} \\ v_{n,k-1} \\ \vdots \\ v_{n,k-L+1} \end{bmatrix} = \begin{bmatrix} d_{n,0} & d_{n,1} & \ldots & d_{n,L-1} \\ 0 & d_{n,0} & \ldots & d_{n,L-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{n,0} \end{bmatrix} \begin{bmatrix} u_{n,k} \\ u_{n,k-1} \\ \vdots \\ u_{n,k-L+1} \end{bmatrix} \tag{4.13}$$

From this equation it is natural to define the vectors $\mathbf{u}_{n,k} = [u_{n,k}, \ldots, u_{n,k-L+1}]^T$ and $\mathbf{v}_{n,k} = [v_{n,k}, \ldots, v_{n,k-L+1}]^T$. Notice that $\mathbf{u}_{n,k}$ is a time-vector for the nth output signal whereas $\mathbf{u}_k$ is the output vector with time-index k. A similar matrix description of the source vector is straightforward. The mixture model and the estimation model are depicted in Figure 4.5.

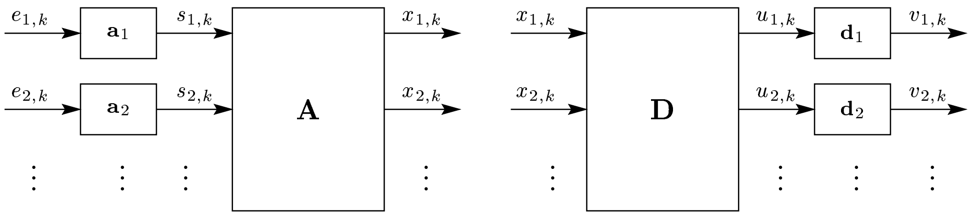


Figure 4.5 Left: The mixture model. $e_{n,k}$ is the nth excitation process, $\mathbf{a}_n$ is the coefficient vector of the nth source, $s_{n,k}$ is the nth source signal, $\mathbf{A}$ is the mixture matrix, and $x_{n,k}$ is the nth observable signal with time-index k. Right: The estimation model. $\mathbf{D}$ is the inverse mixture matrix, $u_{n,k}$ is the nth output signal, $\mathbf{d}_n$ is the filter coefficients of the nth output filter, and $v_{n,k}$ is the nth output signal.

The reward for this cumbersome description of the estimation model is that the following derivation of the likelihood estimator is readily obtained and the derivation is very similar to the derivation in Section 4.1.2.

Because the elements in $\mathbf{v}_{n,k}$ are independent it is straightforward, using Equation 2.13 and Equation 4.13, to obtain the joint pdf. of the nth output vector $\mathbf{u}_{n,k}$

$$f_{\mathbf{u}_{n,k}}(\mathbf{u}) = |d_{n,0}|^L \prod_{l=0}^{L-1} f_{v_{n,k-l}}(v_{n,k-l})$$

To describe the complete probability structure of the output signals the following stacked vector $\mathbf{u} = [\mathbf{u}_{1,k}^T, \ldots, \mathbf{u}_{N,k}^T]^T$ is introduced. Because the output vectors $\mathbf{u}_{n,k}$ are independent, the joint pdf. of $\mathbf{u}$ is the product of the marginal pdf.s

$$f_{\mathbf{u}}(\mathbf{u}) = \prod_{n=1}^{N} f_{\mathbf{u}_{n,k}}(\mathbf{u}) = \prod_{n=1}^{N} |d_{n,0}|^L \prod_{l=0}^{L-1} f_{v_{n,k-l}}(v_{n,k-l})$$

Furthermore, to describe the complete probability structure of the observations, the following stacked observation vector $\mathbf{x} = [\mathbf{x}_{1,k}^T, \ldots, \mathbf{x}_{N,k}^T]^T$ is introduced. The output vector $\mathbf{u}$ is related to the observation vector $\mathbf{x}$ through a block diagonal matrix, i.e. $\mathbf{u} = \text{diag}[\mathbf{D}, \ldots, \mathbf{D}]\,\mathbf{x}$, and thus the pdf. model for the observation vector is

$$f_m(\mathbf{x}) = |\det \mathbf{D}|^L f_{\mathbf{u}}(\mathbf{u}) = |\det \mathbf{D}|^L \prod_{n=1}^{N} |d_{n,0}|^L \prod_{l=0}^{L-1} f_{v_{n,k-l}}(v_{n,k-l}) \tag{4.14}$$

This equation corresponds to Equation 4.2 and for $L = 1$ the equations are identical. The pdf. model in Equation 4.14 is parameterized with the parameters of the estimation model and the log-likelihood function of the parameters is

$$\mathcal{L}(\mathbf{D}, \mathbf{d}) = L \ln|\det \mathbf{D}| + \mathrm{E}\left(L \sum_{n=1}^{N} \ln|d_{n,0}| + \sum_{n=1}^{N} \sum_{l=0}^{L-1} \ln f_{v_{n,k-l}}(v_{n,k-l}) \right) \tag{4.15}$$

where $\mathbf{d} = [\mathbf{d}_1 \ldots \mathbf{d}_N]$ contains the coefficients of the output filters.

To develop an estimator of the parameters in the estimation model, the derivatives of the log-likelihood function with respect to the parameters are determined

$$\frac{\partial}{\partial D_{mn}} \mathcal{L}(\mathbf{D}, \mathbf{d}) = L \left(\mathbf{D}^T\right)^{-1}_{mn} + \mathrm{E}\left(\sum_{l=0}^{L-1} g(v_{m,k-l}) \sum_{q=0}^{L-1} d_{m,q} x_{n,k-l-q} \right) \tag{4.16}$$

$$\frac{\partial}{\partial d_{n,l}} \mathcal{L}(\mathbf{D}, \mathbf{d}) = \delta_l \frac{L}{d_{n,0}} + \mathrm{E}\left(\sum_{q=0}^{L-1} g(v_{n,k-q}) u_{n,k-l-q} \right) \tag{4.17}$$

where $g(v_{m,k-l}) = \frac{\partial}{\partial v_{m,k-l}} f_{v_{m,k-l}}(v_{m,k-l})$, and $\left(\mathbf{D}_0^T\right)^{-1}_{mn}$ denote the (m,n)th element of $\left(\mathbf{D}_0^T\right)^{-1}$.

Adaptation Equations Using the derivative of the likelihood function with respect to $\mathbf{D}$ in Equation 4.16 it is straightforward to develop stochastic gradient update equations. The update equation using the relative gradient of the $\mathbf{D}$ matrix can be written on the form

$$\Delta\mathbf{D} = \mathbf{D} + \frac{1}{L}\mathbf{G}\mathbf{D}^T\mathbf{D} \tag{4.18}$$

where $\mathbf{G}$ is a matrix with the mnth element $\mathrm{E}\left(g(\mathbf{v}_m^T)\mathcal{H}\left(\mathbf{x}_{n,k}, \mathbf{x}_{n,k-L+1}\right)\mathbf{d}_m\right)$ and $\mathcal{H}\left(\mathbf{x}_{n,k}, \mathbf{x}_{n,k-L+1}\right)$ is a Hankel matrix with first column $\mathbf{x}_{n,k}$ and last row $\mathbf{x}_{n,k-L+1}$. Similarly the update equations for the output filter are developed from the derivative of the likelihood function with respect to $d_{n,l}$ given in Equation 4.17. The update equations of the output filters using the ordinary gradient are

$$\Delta d_{n,0} = \frac{L}{d_{n,0}} + \mathrm{E}\left(g(\mathbf{v}_{n,k})\mathbf{u}_{n,k-l}\right) \tag{4.19}$$

$$\Delta d_{n,l} = \mathrm{E}\left(g(\mathbf{v}_{n,k})\mathbf{u}_{n,k-l}\right) \quad \text{for} \quad l = 1, \ldots, L-1 \tag{4.20}$$

Similar update equations can be found in [6].

Simulation Example This simulation example considers an instantaneous mixture of two heavy tail source signals, where the source signals are generated as AR(1)-processes with iid. SαS excitation signals. The estimation is based on 1000 samples using the update equations in Equation 4.18, 4.19 and

4.20, and the output filter length is $L = 10$. The expectation in the update equations is estimated over all 1000 samples, and the non-linearity is the sign-function. The mixture matrix and the AR-coefficients are randomly generated. The adaptation constant is $\mu = 10^{-3}$ and the update is using a simple line search where the adaptation coefficient is multiplied with 2 for every line search. The stop criterion is based on the empirical log-likelihood function, and the estimation is stopped when the relative increase in the log-likelihood function is less than 10^{-4}.

In Figure 4.6 the results of a simulation with this simulation setup are shown. For this specific simulation the coefficients of the AR-process are $a_{1,1} = -0.8890$ and $a_{2,1} = -0.7130$ and the excitation signal is an iid. $\mathrm{S}\alpha\mathrm{S}(1.6, 0.5)$-process. The algorithm converged after 22 iterations, and the practical experience is that for this simulation setup convergence is in general obtained in less than 30 iterations. The mixture matrix, the estimation matrix, and the product between those two matrices are

$$\mathbf{A} = \begin{bmatrix} -1.1465 & 1.1892 \\ 1.1909 & -0.0376 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0.0456 & 0.7770 \\ 0.8069 & 0.7428 \end{bmatrix}, \quad \mathbf{DA} = \begin{bmatrix} 0.8731 & 0.0249 \\ -0.0405 & 0.9316 \end{bmatrix}$$

Observe that $\mathbf{DA}$ is almost diagonal. In Figure 4.6(a), 4.6(b) and 4.6(c) the scatter plot of the source signal, the observable signal, and the output signal are shown. Finally in Figure 4.7 the empirical log-likelihood function versus the iteration number is shown.

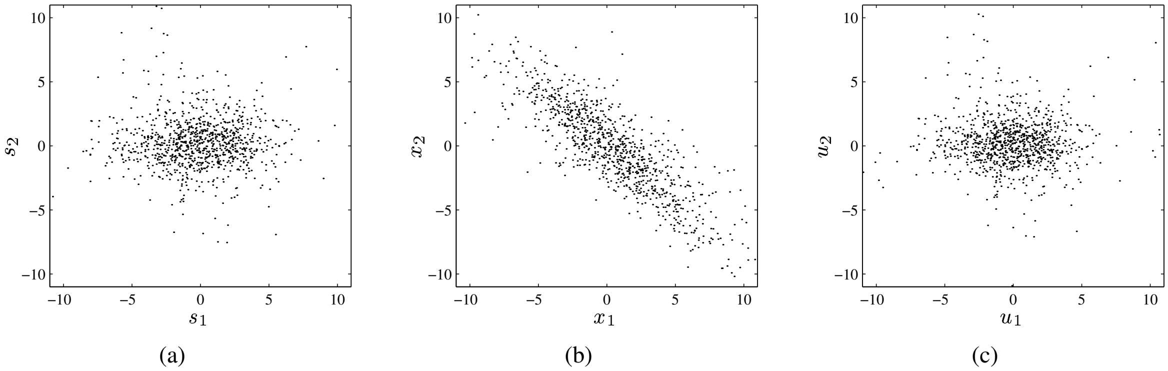


Figure 4.6 (a) Scatter plot of source signal, s_2 versus s_1. (b) Scatter plot of observation signal, x_2 versus x_1. (c) Scatter plot of output signal, u_2 versus u_1.

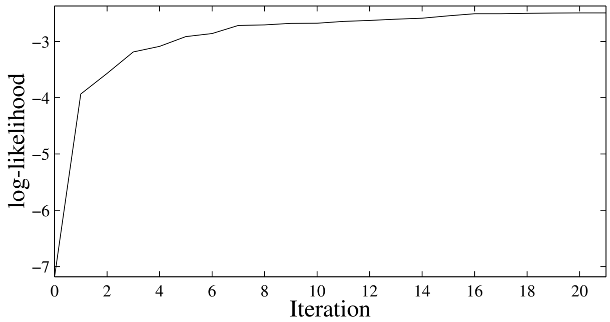


Figure 4.7 Empirical log-likelihood function versus the iteration number.

4.1.4 Convolutive Mixtures of Temporal Independent Sources

The maximum likelihood method for blind separation of convolutive mixtures of temporal independent sources is due to [6]. In the original work three methods, a time-domain, a frequency domain and a hybrid between the time- and frequency-domain, for the separation of convolutive mixtures are derived. The frequency-domain and the hybrid methods have the best convergence properties, however in this work only the time-domain method is considered.

As for the methods in Section 4.1.2 and 4.1.3 it is convenient to establish both a generative model and an estimation model. Let $s_{n,k}$, $x_{n,k}$ and $u_{n,k}$ denote the nth source signal, observable signal and output signal respectively with time index k. The mixture model is a convolutive mixture of the source signals. The estimation model is a linear filtering of the observable signals, and the mth output signal with time index k is

$$u_{m,k} = \sum_{n=1}^{N} \sum_{l=0}^{L-1} D_{mn,l} x_{n,k-l}$$

where $D_{mn,l}$ is the lth filter coefficient from observable signal n to output signal m. For the development of a pdf. model of the observable signals the following matrix notation is introduced. Let the vectors $\mathbf{s}_k = [s_{1,k}, \dots, s_{N,k}]^T$, $\mathbf{x}_k = [x_{1,k}, \dots, x_{N,k}]^T$ and $\mathbf{u}_k = [u_{1,k}, \dots, u_{N,k}]^T$ be the source vector of independent source signals, the observable vector and the output vector of the estimation model respectively with time index k. Furthermore let $D_{mn,l}$ denote the (m,n)th element of the matrix $\mathbf{D}_l$. The estimation model can now be written as

$$\mathbf{u}_k = \sum_{l=0}^{L-1} \mathbf{D}_l \mathbf{x}_{k-l} \tag{4.21}$$

Organizing the $\mathbf{D}_l$ matrices in a block upper triangular form, the complete estimation model can be written as

$$\begin{bmatrix} \mathbf{u}_k \\ \mathbf{u}_{k-1} \\ \vdots \\ \mathbf{u}_{k-L+1} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \dots & \mathbf{D}_{L-1} \\ 0 & \mathbf{D}_0 & \dots & \mathbf{D}_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{D}_0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k-1} \\ \vdots \\ \mathbf{x}_{k-L+1} \end{bmatrix} \tag{4.22}$$

The determinant of the block upper diagonal matrix is the product of the determinants of the matrices in the diagonal. It is natural to define the vectors $\mathbf{x} = [\mathbf{x}_k^T, \dots, \mathbf{x}_{k-L+1}^T]^T$ and $\mathbf{u} = [\mathbf{u}_k^T, \dots, \mathbf{u}_{k-L+1}^T]^T$, and from Equation 4.22 the pdf. of the stacked vector $\mathbf{x}$ is readily obtained

$$f_m(\mathbf{x}) = |\det \mathbf{D}_0|^L \prod_{l=0}^{L-1} f_{\mathbf{u}_{k-l}}(\mathbf{u}_{k-l}) = |\det \mathbf{D}_0|^L \prod_{l=0}^{L-1} \prod_{n=1}^{N} f_{u_{n,k-l}}(u_{n,k-l}) \tag{4.23}$$

Since the vectors $\mathbf{u}_k$ are independent the pdf. of $\mathbf{u}$ is the product of the marginal pdf.s of $\mathbf{u}_k$, and thereby the first identity is obtained. The second identity is because the elements in $\mathbf{u}_k$ are independent and therefore the pdf. of $\mathbf{u}_k$ is the product of the marginal pdf.s.

It is interesting to notice that Equation 4.23 is very similar to Equation 4.2 and Equation 4.14, and for $L = 1$ Equation 4.2, 4.14 and 4.23 are identical. The pdf. model in Equation 4.23 is parameterized with the coefficients of the estimation model and the log-likelihood function of the estimation model is

$$\mathcal{L}(\mathbf{D}) = L \ln |\det \mathbf{D}_0| + \mathrm{E}\left(\sum_{l=0}^{L-1} \sum_{n=1}^{N} \ln f_{u_{n,k-l}}(u_{n,k-l}) \right) \tag{4.24}$$

To develop an estimator the derivative of the log-likelihood function is determined

$$\begin{aligned}\frac{\partial}{\partial D_{mn,0}}\mathcal{L}(\mathbf{D}) &= L\left(\mathbf{D}_0^T\right)^{-1}_{mn} + \mathrm{E}\left(\sum_{l=0}^{L-1} g(u_{m,k-l})x_{n,k-l}\right) &\quad (4.25)\\ \frac{\partial}{\partial D_{mn,q}}\mathcal{L}(\mathbf{D}) &= \mathrm{E}\left(\sum_{l=0}^{L-1} g(u_{m,k-l})x_{n,k-l-q}\right) \quad \text{for} \quad q = 1,\dots,L-1 &\quad (4.26)\end{aligned}$$

where $g(u_{m,k-l}) = \frac{\partial}{\partial u_{m,k-l}} f_{u_{m,k-l}}(u_{m,k-l})$ and $\left(\mathbf{D}_0^T\right)^{-1}_{mn}$ denote the (m,n)th element of $\left(\mathbf{D}_0^T\right)^{-1}$.

Adaptation Equations Using the gradient of the likelihood in Equation 4.25 and 4.26 the stochastic gradient algorithm, using the ordinary gradient with averaging over K samples has the update equations

$$\begin{aligned}\Delta\mathbf{D}_0 &= \left(\mathbf{D}_0^T\right)^{-1} + \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=0}^{L-1} g(\mathbf{u}_{k-l})\mathbf{x}_{k-l}^T &\quad (4.27)\\ \Delta\mathbf{D}_q &= \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=0}^{L-1} g(\mathbf{u}_{k-l})\mathbf{x}_{k-l-q}^T \quad \text{for} \quad q = 1,\dots,L-1 &\quad (4.28)\end{aligned}$$

The relative gradient update equation is obtained by right multiplying Equation 4.27 and Equation 4.28 with $\mathbf{D}_l^T\mathbf{D}_l$. Similar update equations can be found in [6].

Simulation Example Consider a convolutive mixture of two iid. source signals with normalized double exponential pdf., and with convolution filters of length 20. The estimation model is equivalent to the generative model, that is $\mathbf{A}_l$ and $\mathbf{D}_l \in \mathbb{R}^{2\times 2}$ for $l = 1,\dots,20$. The estimation of $\mathbf{D}_l$ is over 2000 samples, the gradient is the relative gradient and the averaging is over all 2000 samples. The adaptation constant is $\mu = 10^{-3}$ and the update is using a simple line search where the adaptation coefficient is multiplied with 2 for every line search. The stop criterion is based on the empirical log-likelihood function, and the estimation is stopped when the relative increase in the log-likelihood function is less than 10^{-4}.

The coefficients of the convolutive filter are generated as iid. Gaussian random variables with exponentially decreasing variance, and in Figure 4.8 the impulse responses of the convolutive filters for this specific simulation example are depicted. There is no guarantee for convergence to the global maximum. The practical experience with the algorithm is that convergence is relative slow and the estimation will in general result in a local maximum. The shown simulation example is a very typical result, and in general reasonable solutions are obtained.

In Figure 4.9 the impulse responses of the estimated deconvolution filters for this simulation example are depicted. The overall impulse response is the convolution of the convolution filters and the deconvolution filters which are shown in Figure 4.10. It appears that the overall impulse responses are close to an impulse in the anti-diagonal and have small gain in the diagonal. Thus there is a channel permutation, but apart from that the estimated deconvolution filters are close to the inverse of the convolution filters.

Figure 4.11(a), 4.11(b) and 4.11(c) show the scatter plots of the source signals, the observable signals, and the estimated signal respectively. From the scatter plot of the estimated signals it seems that the estimated signals are independent.

The empirical log-likelihood function versus the iteration number is depicted in Figure 4.12. In this simulation convergence was obtained after 363 iterations.

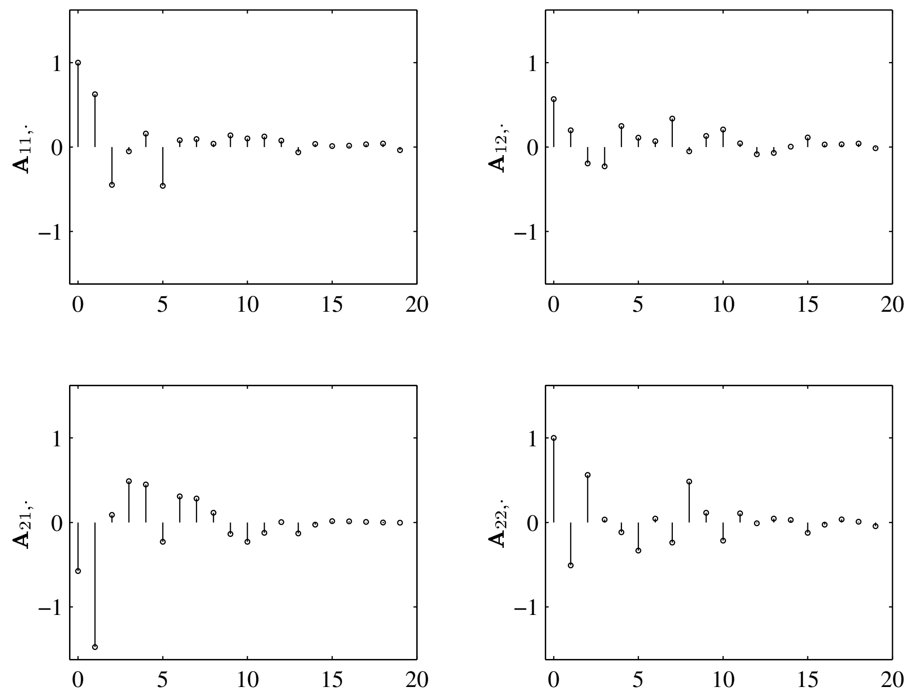


Figure 4.8 Impulse responses for the convolutive filters. $\mathbf{A}_{mn,\cdot}$ denote the coefficients in the filter from input s_n to observation x_m.

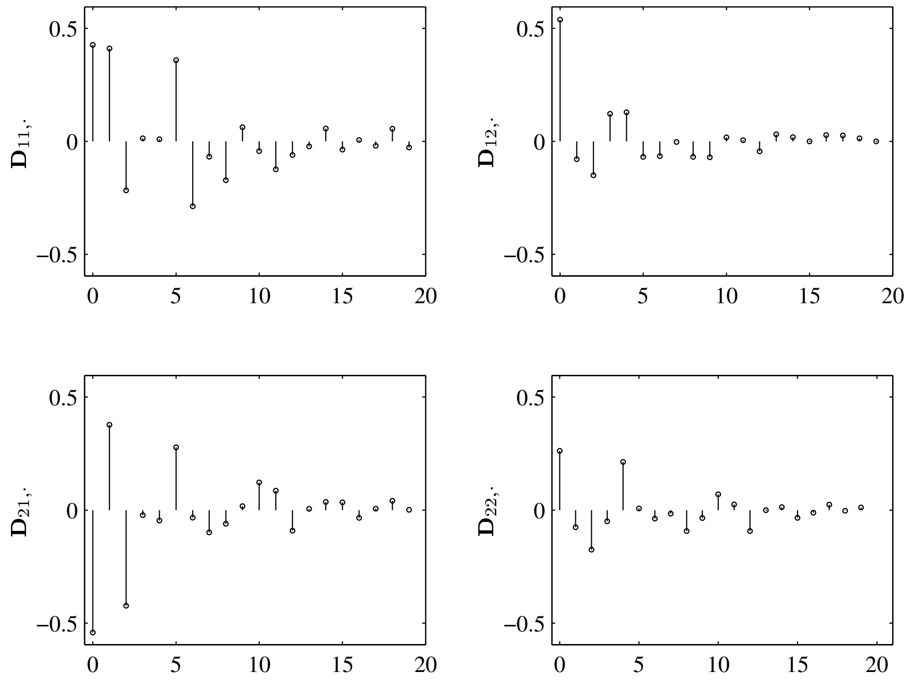


Figure 4.9 Impulse responses for the deconvolution filters. $\mathbf{D}_{mn,\cdot}$ denote the coefficients in the filter from observation x_n to output u_m.

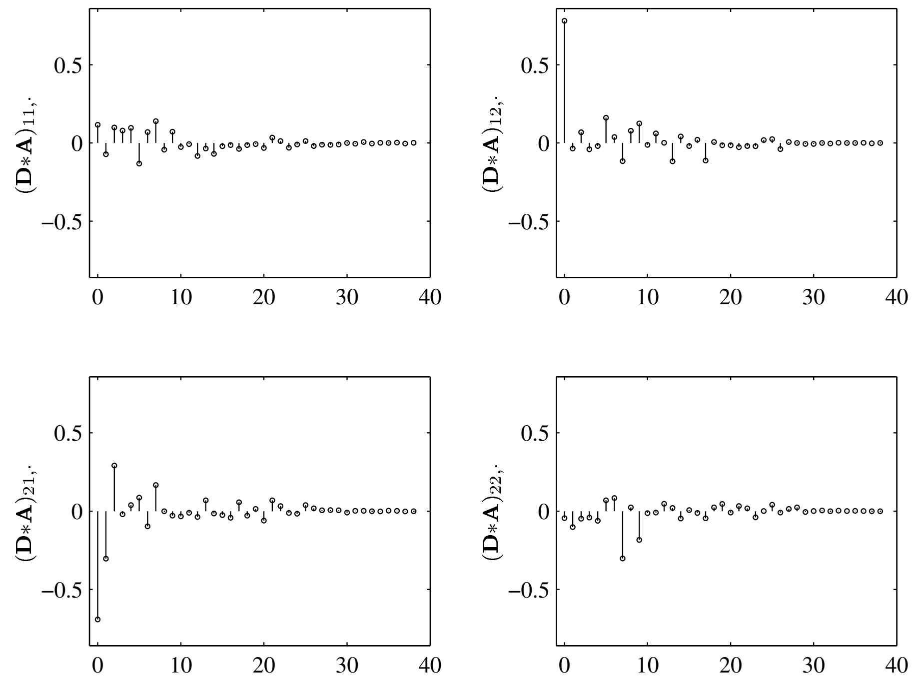


Figure 4.10 Impulse responses for the mixing filters convolved with the separating filters. $(\mathbf{D} * \mathbf{A})_{mn,\cdot}$ denote the overall impulse response from input s_n to output u_n.

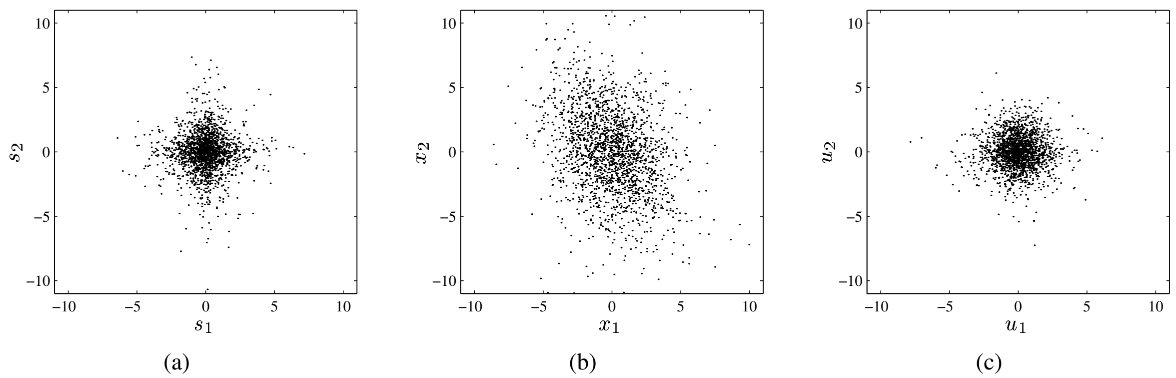


Figure 4.11 (a) Scatter plot of source signal, s_2 versus s_1. (b) Scatter plot of observation signal, x_2 versus x_1. (c) Scatter plot of output signal, u_2 versus u_1.

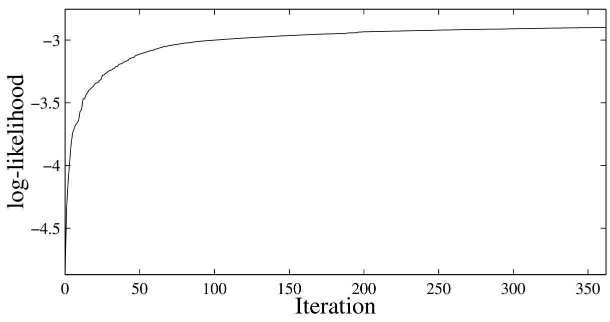


Figure 4.12 Empirical log-likelihood function versus the iteration number.

4.2 Methods Based on Joint Diagonalization

This section considers methods for blind source separation based on joint diagonalization of estimated matrices. The problem of joint diagonalization using non-orthogonal matrices is a general problem, and in blind signal separation several methods based on joint diagonalization have been proposed, see [69, 94, 24, 11, 88, 1].

The general joint diagonalization problem can be formulated as follows. Based on a set $\mathcal{M}$ of L matrices $\mathbf{M}_1, \ldots, \mathbf{M}_L \in \mathbb{C}^{N\times N}$ we seek a non-orthogonal coefficient matrix, $\mathbf{A}$, that yields the exact joint diagonalization

$$\mathbf{M}_l = \mathbf{A}\mathbf{\Lambda}_l\mathbf{A}^H \quad \text{for} \quad l = 1, \ldots, L \tag{4.29}$$

where $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_K$ are diagonal matrices corresponding to the set $\mathcal{M}$. In general the set $\mathcal{M}$ is not available, and the joint diagonalization is based on an estimated set $\hat{\mathcal{M}}$ of $\hat{\mathbf{M}}_1, \ldots, \hat{\mathbf{M}}_L$. However, due to estimation errors no exact diagonalization of $\hat{\mathcal{M}}$ exists, and in practice the solution is to find a coefficient matrix, $\hat{\mathbf{A}}$, that yields an approximated joint diagonalization.

Joint Diagonalization Methods The simplest approach for solving the joint diagonalization problem is to restrict $\mathcal{M}$ to a set of two matrices, and formulate the eigenvalue problem

$$\mathbf{M}_1\mathbf{M}_2^{-1}\mathbf{A} = \mathbf{A}\mathbf{\Lambda}_1\mathbf{\Lambda}_2^{-1} \tag{4.30}$$

This approach may be attractive if $\mathbf{M}_1$ and $\mathbf{M}_2$ are a well conditioned constellation, and if $\mathbf{M}_1$ and $\mathbf{M}_2$ can be accurately estimated. It is interesting to notice that contrary to the methods in Section 4.1 a method based on joint diagonalization of two matrices has an explicit solution for the mixture matrix.

Several algorithms for solving the approximate joint diagonalization problem for $L > 2$ are proposed in the literature. The method used in this section is proposed in [93]. The method is iterative and is minimizing the weighted least-squares criterion

$$\sum_{l=1}^{L} \|\hat{\mathbf{M}}_l - \hat{\mathbf{A}}\hat{\mathbf{\Lambda}}_l\hat{\mathbf{A}}^H\|_F^2 w_l$$

where w_l is the weight assigned to $\hat{\mathbf{M}}_l$, $\|\cdot\|_F$ is the Frobenius norm and $\hat{\mathbf{\Lambda}}_l$ is the estimated diagonal matrix corresponding to $\hat{\mathbf{M}}_l$. Other methods are proposed in [24, 90, 25].

4.2.1 Signal Separation Based on Time-Delayed Decorrelation

In [69] Molgedey and Schuster proposed a signal separation method based on time-delayed decorrelation of the observable signals. The generative model is an instantaneous mixture of stationary, temporal correlated, mutual independent source signals. The presentation in the framework of joint diagonalization is more general than in [69].

From the generative model the observation vector with time-index k is

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the mixture and $\mathbf{s}_k = [s_{1,k} \ldots s_{N,k}]^T$ is the source vector. The sources are assumed to be mutually independent, zero mean, stationary linear stochastic processes, with different autocorrelation functions. Define the set $\mathcal{M}$ containing the spatial covariance matrices of the observation vector

$$\begin{aligned}
\mathbf{R}_{\mathbf{x},l} &= \mathrm{E}\left(\mathbf{x}_k\mathbf{x}_{k+l}^T\right) \\
&= \mathrm{E}\left(\mathbf{A}\mathbf{s}_k\mathbf{s}_{k+l}^T\mathbf{A}^T\right) = \mathbf{A}\mathrm{E}\left(\mathbf{s}_k\mathbf{s}_{k+1}^T\right)\mathbf{A}^T = \mathbf{A}\mathbf{R}_{\mathbf{s},l}\mathbf{A}^T
\end{aligned} \tag{4.31}$$

for L different l. Since the sources are independent the spatial covariance matrix of the sources, $\mathbf{R}_{\mathbf{s},l}$, is diagonal, and this leads to the joint diagonalization problem in 4.29. In the original form the algorithm is based on two spatial covariance matrices, one with time shift l and one with time shift 0. Thus $\mathcal{M}$ contains $\mathbf{R}_{\mathbf{x},0}$ and $\mathbf{R}_{\mathbf{x},l}$, and the diagonalization is obtained by solving the eigenvalue problem from Equation 4.30

$$\mathbf{R}_{\mathbf{x},l}\mathbf{R}_{\mathbf{x},0}^{-1}\mathbf{A} = \mathbf{A}\mathbf{R}_{\mathbf{s},l}\mathbf{R}_{\mathbf{s},0}^{-1} \tag{4.32}$$

If the second moment of the sources exist, and if $\mathrm{E}\left(s_{n,k}s_{n,k+l}\right)/\mathrm{E}\left(s_{n,k}s_{n,k}\right)$, $n = 1, \ldots, N$ is a set of distinct numbers, then the eigenvalue problem has a solution.

In practice the set of L spatial covariance matrices must be estimated from the data. The mixture matrix is thus obtained from an approximated joint diagonalization of the set of estimated spatial covariance matrices.

Spatial Covariation Matrices for SαS Sources For non-Gaussian stable processes the second order moment does not exist, and consequently also the spatial covariance matrices from the previous do not exist. Fortunately the method generalizes to spatial covariation matrices. This result is published in [35] included in Appendix A.3. The generative model is an instantaneous mixture of stationary, mutually independent, linear stable processes.

First we need to generalize the linearity properties for the covariation from Section 3.3.2. Let $\mathbf{x} = [x_1 \ldots x_M]^T$ and $\mathbf{y} = [y_1 \ldots y_N]^T$ be jointly SαS random vectors where the elements in $\mathbf{y}$ are independent. Let $\mathbf{a} \in \mathbb{R}^M$ and $\mathbf{b} \in \mathbb{R}^N$ be constant vectors, then

$$\left[\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}\right]_\alpha = \mathbf{a}^T\left[\mathbf{x}, \mathbf{y}^T\right]_\alpha \mathbf{b}^{T^{\langle\alpha-1\rangle}} = \mathbf{a}^T\mathbf{C}\mathbf{b}^{T^{\langle\alpha-1\rangle}} \tag{4.33}$$

where $\mathbf{C} = \left[\mathbf{x}, \mathbf{y}^T\right]_\alpha$ is the $M \times N$ covariation matrix with elements $c_{m,n} = [x_m, y_n]_\alpha$, and $\mathbf{b}^{T^{\langle p\rangle}}$ is the transpose of the element-wise signed power p of the vector $\mathbf{b}$.

Consider again the generative model with the instantaneous mixing matrix $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_M]^T$, where $\mathbf{a}_m^T$ is the mth row of $\mathbf{A}$. Let $x_{m,k} = \mathbf{a}_m^T\mathbf{s}_k$ denote the mth element of the observable vector with time-index k. Using the linearity property for the covariation, the covariation between the mth and the time-shifted nth element of the vector $\mathbf{x}$ is

$$[x_{m,k}, x_{n,k+l}]_\alpha = \left[\mathbf{a}_m^T\mathbf{s}_k, \mathbf{a}_n^T\mathbf{s}_{k+l}\right]_\alpha = \mathbf{a}_m^T\left[\mathbf{s}_k, \mathbf{s}_{k+l}^T\right]_\alpha \mathbf{a}_n^{\langle\alpha-1\rangle} = \mathbf{a}_m^T\mathbf{C}_{\mathbf{s},l}\mathbf{a}_n^{\langle\alpha-1\rangle} \tag{4.34}$$

where $\mathbf{C}_{\mathbf{s},l} = \left[\mathbf{s}_l, \mathbf{s}_{k+l}^T\right]_\alpha$. It is natural to refer to $\mathbf{C}_{\mathbf{s},l}$ as the time-shifted spatial covariation matrix of the source vector $\mathbf{s}$. Because the source signals are independent $\mathbf{C}_{\mathbf{s},l}$ is a diagonal matrix. Now define the time shifted spatial covariation matrix of the observable vector $\mathbf{x}$

$$\mathbf{C}_{\mathbf{x},l} = \left[\mathbf{x}_l, \mathbf{x}_{k+l}^T\right]_\alpha$$

The (m,n)th element of $\mathbf{C}_{\mathbf{x},l}$ is $[x_{m,k}, x_{n,k+l}]_\alpha$ given in Equation 4.34. From this it appears that the spatial covariation matrix can be written as

$$\mathbf{C}_{\mathbf{x},l} = \mathbf{A}\mathbf{C}_{\mathbf{s},l}\mathbf{A}^{T^{\langle\alpha-1\rangle}} \tag{4.35}$$

Notice that this relation between the spatial covariation matrices is very similar to the relation between the spatial covariance matrices in Equation 4.31. Furthermore Equation 4.35 is recognized as a slightly modified version of the general joint diagonalization problem from Equation 4.29. Applying the eigenvalue problem approach from Equation 4.30, the mixing matrix is the solution to

$$\mathbf{C}_{\mathbf{x},l_1}\mathbf{C}_{\mathbf{x},l_2}^{-1}\mathbf{A} = \mathbf{A}\mathbf{C}_{\mathbf{s},l_1}\mathbf{C}_{\mathbf{s},l_2}^{-1}$$

where $\mathbf{C}_{\mathbf{x},l_1}$ and $\mathbf{C}_{\mathbf{x},l_1}$ is the spatial covariation matrix to two different time shifts.

In practice the spatial covariation matrices must be estimated. Sample estimates of the spatial covariation matrices are easily obtained based on fractional lower order moments. For simulation examples see Appendix A.3.

In Section 3.3.1 the normalized correlation function is introduced. Considering the right hand side of the eigenvalue problem in Equation 4.32, it is seen that $\mathbf{R}_{\mathbf{s},l}\mathbf{R}_{\mathbf{s},0}^{-1}$ is simply the normalized spatial covariation matrix, and this matrix is well defined even in the case of stable processes. Since the right hand side is well defined so is the left hand side and the method based on delayed decorrelation is thus valid even in the case of stable processes. This was conjectured in [35], but the theoretical justification was not established.

4.2.2 Blind Signal Separation based on the Generalized Cumulant Generating Function

This method is due to Yeredor [94] and the generative model is an instantaneous mixture of independent source signals. The observation model is $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$, where $\mathbf{s}_k$ is the source vector with time-index k, and $\mathbf{x}_k$ is the corresponding observation vector. The method is based on a joint diagonalization of matrices of the second derivative of the generalized cumulant generating function evaluated for different values of the characteristic function variable.

The moment generation function is defined in Equation 2.2 for $t \in \mathbb{R}$. For $\mathbf{t} \in \mathbb{C}^N$ the function is termed the generalized characteristic function (g.chf.), i.e. the generalized characteristic function of the source vector is $\phi_{\mathbf{s}}(\mathbf{t}) = \mathrm{E}\left(\exp(\mathbf{t}^T\mathbf{s})\right)$ where $\mathbf{t} \in \mathbb{C}^N$. The chf. exists for all $\mathbf{t} \in \mathbb{R}^N$ because any bounded function is integrable with respect to any distribution function. For source vectors with finite support distributions the g.chf. exists for all $\mathbf{t} \in \mathbb{C}^N$; for source vectors with infinite support distributions the existence is in general only assured for $\mathbf{t}$ with non-zero imaginary part. If the elements in the source vector are statistically independent then the g.chf. is a product over the marginal g.chf.s

$$\phi_{\mathbf{s}}(\mathbf{t}) = \prod_{n=1}^{N} \phi_{s_n}(t_n) \tag{4.36}$$

where $\phi_{s_n}(t_n) = \mathrm{E}\left(\exp(t_n s_n)\right)$. From the g.chf. the generalized cumulant generating function (g.cgf.) is defined as $\psi_{\mathbf{s}}(\mathbf{t}) = \ln \phi_{\mathbf{s}}(\mathbf{t})$. The g.cgf. of the source vector is a sum over the marginal g.cgf.s

$$\psi_{\mathbf{s}}(\mathbf{t}) = \sum_{n=1}^{N} \psi_{s_n}(t_n) \tag{4.37}$$

where $\psi_{s_n}(t_n) = \ln\left(\phi_{s_n}(t_n)\right)$. Define a matrix $\mathbf{\Psi}_{\mathbf{s}}(\mathbf{t})$ as the second derivative of $\psi_{\mathbf{s}}(\mathbf{t})$ with respect to $\mathbf{t}$

$$\mathbf{\Psi}_{\mathbf{s}}(\mathbf{t}) \triangleq \frac{\partial}{\partial \mathbf{t}} \left(\frac{\partial}{\partial \mathbf{t}} \psi_{\mathbf{s}}(\mathbf{t}) \right)^T$$

From the sum-form of the g.cgf. in Equation 4.37 it is seen that $\mathbf{\Psi}_{\mathbf{s}}(\mathbf{t})$ is a diagonal matrix

$$\mathbf{\Psi}_{\mathbf{s}}(\mathbf{t}) = \mathrm{diag}\left[\frac{\partial^2 \psi_{s_1}(t_1)}{\partial t_1^2} \cdots \frac{\partial^2 \psi_{s_N}(t_N)}{\partial t_N^2} \right]$$

The relations between the g.chf. for the observation vector and the source vector are $\psi_{\mathbf{x}}(\mathbf{t}) = \psi_{\mathbf{s}}(\mathbf{A}^T\mathbf{t})$, and similarly for the second g.cgf. $\psi_{\mathbf{x}}(\mathbf{t}) = \psi_{\mathbf{x}}(\mathbf{A}^T\mathbf{t})$. The second derivative of the g.cgf. for the observation vector is

$$\mathbf{\Psi}_{\mathbf{x}}(\mathbf{t}) = \frac{\partial}{\partial \mathbf{t}} \left(\frac{\partial}{\partial \mathbf{t}} \psi_{\mathbf{x}}(\mathbf{t}) \right)^T = \frac{\partial}{\partial \mathbf{t}} \left(\frac{\partial}{\partial \mathbf{t}} \psi_{\mathbf{s}}(\mathbf{A}^T\mathbf{t}) \right)^T = \mathbf{A}\mathbf{\Psi}_{\mathbf{s}}(\mathbf{A}^T\mathbf{t})\mathbf{A}^T \tag{4.38}$$

Let $\hat{\mathcal{M}}$ be the set of $\mathbf{\Psi_x}(\mathbf{t}_1), \ldots, \mathbf{\Psi_x}(\mathbf{t}_L)$ for distinct evaluation points $\mathbf{t}_l$, $l = 1, \ldots, L$. Equation 4.38 is then recognized as a joint diagonalization problem of the form of Equation 4.29. The mixing matrix, $\mathbf{A}$, is thus obtained as a joint diagonalization of $\hat{\mathcal{M}}$.

It is interesting to notice that $\mathbf{\Psi_s}(\mathbf{0})$ and $\mathbf{\Psi_x}(\mathbf{0})$ are the spatial covariance matrix for the source and observation vector respectively. In that case Equation 4.38 is identical to Equation 4.31.

Existence of a Solution For a discussion of the identifiability of the mixture consider the $N \times N$ real mixture $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_N]$, where $\mathbf{a}_n$ is the nth column of $\mathbf{A}$, and the set $\hat{\mathcal{M}}$ of two matrices $\mathbf{\Psi_x}(\mathbf{t}_1)$, $\mathbf{\Psi_x}(\mathbf{t}_2)$. The joint diagonalization can then be formulated as the eigenvalue problem

$$\mathbf{\Psi_x}(\mathbf{t}_1)\mathbf{\Psi_x}^{-1}(\mathbf{t}_2)\mathbf{A} = \mathbf{A}\mathbf{\Psi_s}(\mathbf{A}^T\mathbf{t}_1)\mathbf{\Psi_s}^{-1}(\mathbf{A}^T\mathbf{t}_2) \tag{4.39}$$

Define the scalars $t_{n,1} = \mathbf{a}_n^T\mathbf{t}_1$ and $t_{n,2} = \mathbf{a}_n^T\mathbf{t}_2$. If the conditions

(C1) $$\frac{\partial^2\psi_{s_n}(t_{n,1})}{\partial(t_{n,1})^2} \quad \text{and} \quad \frac{\partial^2\psi_{s_n}(t_{n,2})}{\partial(t_{n,2})^2} \quad \text{exist} \quad \forall\, n = 1, \ldots, N$$

(C2) $$\frac{\partial^2\psi_{s_n}(t_{n,2})}{\partial(t_{n,2})^2} \neq 0 \quad \forall\, n = 1, \ldots, N$$

(C3) $$\frac{\partial^2\psi_{s_n}(t_{n,1})}{\partial(t_{n,1})^2}\left(\frac{\partial^2\psi_{s_n}(t_{n,2})}{\partial(t_{n,2})^2}\right)^{-1}, \quad n = 1, \ldots, N \quad \text{is a set of distinct numbers}$$

hold then the matrices $\mathbf{\Psi_x}(\mathbf{t}_1)$, $\mathbf{\Psi_x}(\mathbf{t}_2)$ exist, and the eigenvalue problem in Equation 4.39 has a solution. Thus under condition C1-C3 the mixture $\mathbf{A}$ can be identified up to a scale and a permutation.

For a Gaussian distributed signal s_n we have $\partial^2\psi_{s_n}(t)/\partial t^2 = \sigma_n^2$ where σ_n^2 is the variance of source s_n. i.e. the second derivative of the cgf. is independent of the evaluation point t. Consequently condition C3 requires that at most one of the signals must be Gaussian (this is a well-known result [22]). For all non-Gaussian distributions where $\partial^2\psi_{s_n}(t)/\partial t^2$ exists, $\partial^2\psi_{s_n}(t)/\partial t^2$ is a (non-constant) function of t, and consequently condition C3 is not very restrictive.

Estimation of $\mathbf{\Psi_x}(\mathbf{t})$ In practice the set of L matrices $\mathbf{\Psi_x}(\mathbf{t}_l)$, $l = 1, \ldots, L$ are estimated from samples of the observation vector. The second derivative of the g.cgf. is given by

$$\mathbf{\Psi_x}(\mathbf{t}_l) = \frac{\mathrm{E}\left(\mathbf{x}\mathbf{x}^T\exp(\mathbf{t}^T\mathbf{x})\right)}{\mathrm{E}\left(\exp(\mathbf{t}^T\mathbf{x})\right)} - \frac{\mathrm{E}\left(\mathbf{x}\exp(\mathbf{t}^T\mathbf{x})\right)\mathrm{E}\left(\mathbf{x}^T\exp(\mathbf{t}^T\mathbf{x})\right)}{\mathrm{E}\left(\exp(\mathbf{t}^T\mathbf{x})\right)^2} \tag{4.40}$$

A direct way to obtain estimates of $\mathbf{\Psi_x}(\mathbf{t}_l)$ is simply to substitute the expectation values in Equation 4.40 with sample estimates.

It is interesting to notice that in general if rank$(\mathbf{A}) = N$ and $\mathbf{t}_l \neq \mathbf{0}$ then $\mathbf{\Psi_x}(\mathbf{t}_l)$, $l = 1, \ldots, L$ exists even in the case of SαS vectors. However, the sample properties of $\mathbf{\Psi_x}(\mathbf{t}_l)$ using the above approach are so poor that it is unattractive.

4.2.3 Joint Diagonalization of Cumulant Matrices

In [24][5] Cardoso and Souloumiac proposed a method for blind identification of instantaneous mixtures based on joint diagonalization of cumulant matrices. This presentation deviates from the original presentation in two significant ways. First, the presentation is restricted to real mixture matrices, whereas

[5]This is the original reference for the JADE-algorithm (Joint Approximate Diagonalization of Eigenmatrices).

the original presentation is using complex mixture matrices. The reason for this restriction is that only real valued mixture matrices are considered in the remainder part of this thesis, and the present derivation is easily generalized into complex mixture matrices. Second, the derivation is modified so that the joint diagonalization results in a general matrix, where the original derivation was based on a spatial pre-whitening and then finding a unitary mixture matrix.

The observation model is $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{s}$ is the source vector, $\mathbf{x}$ is the corresponding observation vector, notice that the time index is omitted. Let $\mathbf{A} = [\mathbf{a}_{\cdot 1} \ldots \mathbf{a}_{\cdot N}] = [\mathbf{a}_{1\cdot} \ldots \mathbf{a}_{N\cdot}]^T$, i.e. $\mathbf{a}_{\cdot n}$ is the nth column and $\mathbf{a}_{n\cdot}^T$ is the nth row of the mixture.

Consider the general 4th order cumulant of the observation vector, $\mathrm{cum}\,(x_m, x_n, x_k, x_l)$. Inserting the inner-product form of the elements in the observation vector, the 4th order cumulant is

$$\mathrm{cum}\,(x_m, x_n, x_k, x_l) = \mathrm{cum}\,(\mathbf{a}_{m\cdot}\mathbf{s}, \mathbf{a}_{n\cdot}\mathbf{s}, \mathbf{a}_{k\cdot}\mathbf{s}, \mathbf{a}_{l\cdot}\mathbf{s})$$

Since the sources are independent it is straightforward, using Property 2.1 and Property 2.3 from Section 2.1.3, to establish

$$\mathrm{cum}\,(x_m, x_n, x_k, x_l) = \mathbf{a}_{m\cdot}^T \mathrm{diag}\left[c_{s_1} a_{k1} a_{l1}, \ldots, c_{s_N} a_{kN} a_{lN}\right] \mathbf{a}_{n\cdot} \tag{4.41}$$

where $c_{s_n} = \mathrm{cum}\,(s_n, s_n, s_n, s_n)$, i.e. the 4th cumulant of the nth source distribution. The (k,l)th parallel cumulant slice is a matrix of cumulants where the (m,n)th entry is given by Equation 4.41. Now form a cumulant matrix $\mathbf{Q}(\mathbf{W})$ as a linear combination of (k,l)th parallel cumulant slices, where $\mathbf{W}$ is a real weight matrix. The (k,l)th entry in the weight matrix is w_{kl} which is the weight for the (k,l)th cumulant slice. Thus the (m,n)th element of $\mathbf{Q}(\mathbf{W})$ is

$$q_{mn} = \sum_{k=1}^{N}\sum_{l=1}^{N} \mathrm{cum}\,(x_m, x_n, x_k, x_l)\, w_{kl} \tag{4.42}$$

Using Equation 4.41 the (m,n)th element of the cumulant matrix is

$$\begin{aligned} q_{mn} &= \sum_{k=1}^{N}\sum_{l=1}^{N} \mathbf{a}_{m\cdot}^T \mathrm{diag}\left[c_{s_1} a_{k1} a_{l1}, \ldots, c_{s_N} a_{kN} a_{lN}\right] w_{lk} \mathbf{a}_{n\cdot} \\ &= \mathbf{a}_{m\cdot}^T \mathrm{diag}\left[c_{s_1} \mathbf{a}_{\cdot 1}^T \mathbf{W} \mathbf{a}_{\cdot 1}, \ldots, c_{s_N} \mathbf{a}_{\cdot N}^T \mathbf{W} \mathbf{a}_{\cdot N}\right] \mathbf{a}_{n\cdot} \end{aligned}$$

The cumulant matrix can now be written as

$$\mathbf{Q}(\mathbf{W}) = \mathbf{A}\mathbf{\Lambda}_{\mathbf{W}}\mathbf{A}^T$$

where

$$\mathbf{\Lambda}_{\mathbf{W}} = \mathrm{diag}\left[c_{s_1} \mathbf{a}_{\cdot 1}^T \mathbf{W} \mathbf{a}_{\cdot 1}, \ldots, c_{s_N} \mathbf{a}_{\cdot N}^T \mathbf{W} \mathbf{a}_{\cdot N}\right]$$

This is recognized as a joint diagonalization problem on the form of Equation 4.29. The joint diagonalization is determined from the set $\mathcal{M}$ of L matrices $\mathbf{Q}(\mathbf{W}_l)$, $l = 1, \ldots, L$. A drawback is that there are no general guidelines for choosing the set of weight matrices. Alternatively, the diagonalization can be determined from the complete set of parallel cumulant slice matrices

$$\mathcal{M} = \left\{\mathbf{Q}(\mathbf{1}_k \mathbf{1}_l^T)\right\} \qquad \forall\, k,l = 1, \ldots, N$$

where $\mathbf{1}_k$ denotes a $N \times 1$ vector with 1 in the kth position and zero elsewhere. In practice the joint diagonalization is determined from a set $\hat{\mathcal{M}}$ of estimated $\mathbf{Q}(\mathbf{W}_l)$, $l = 1, \ldots, L$ matrices.

4.2.4 A Unifying Method

It is straightforward to combine the methods considered in Section 4.2.1, 4.2.2 and 4.2.3 into a unified method. Let $\hat{\mathcal{M}}_1$, $\hat{\mathcal{M}}_2$ and $\hat{\mathcal{M}}_3$ denote the three sets of estimated matrices for which the approximate joint diagonalization in Section 4.2.1, 4.2.2 and 4.2.3 is determined. Now construct a new set $\hat{\mathcal{M}}_4$ which is a subset, or the full set, of the matrices in $\hat{\mathcal{M}}_1$, $\hat{\mathcal{M}}_2$ and $\hat{\mathcal{M}}_3$. The unified method is then to determine the approximate joint diagonalization over the set $\hat{\mathcal{M}}_4$.

For combination of sets of matrices in an approximately joint diagonalization problem attention must be paid to the scaling of the matrices to avoid that a subset of the matrices dominates the diagonalization. For the weighted least square method [93] this can be solved by assigning appropriate weights to the matrices.

4.2.5 Comparison of Methods Based on Joint Diagonalization

A single simulation study of the methods in Section 4.2.1, 4.2.2, 4.2.3 and 4.2.4 is provided. A thorough and systematic comparison, theoretical as well as experimental, showing the characteristics of the methods has not been done. The simulation study exhibits typical characteristics, but the results must be interpreted with attention to the simulation setup.

Let $\mathbf{A}$ be the mixture matrix, $\mathbf{D}$ be the unpermuted estimated diagonalization matrix, and $\mathbf{a}_{\cdot n}$ and $\mathbf{d}_{\cdot n}$ be the normalized nth column of $\mathbf{A}$ and $\mathbf{D}$ respectively. To compare the methods the following performance function is proposed

$$\Upsilon = \frac{1}{N} \sum_{n=1}^{N} |\mathbf{d}_{\cdot n}^T \mathbf{a}_{\cdot n}| \tag{4.43}$$

Since $\|\mathbf{a}_{\cdot n}\|_2 = 1$ and $\|\mathbf{d}_{\cdot n}\|_2 = 1$ the inner product $\mathbf{d}_{\cdot n}^T \mathbf{a}_{\cdot n} = \cos\theta$ where θ is the angle between $\mathbf{a}_{\cdot n}$ and $\mathbf{d}_{\cdot n}$. The performance function is a measure of how close the estimated diagonalization matrix is to the mixture matrix. The following simulation setup satisfies the conditions for the methods in Section 4.2.1, 4.2.2 and 4.2.3.

Simulation Setup The generative model is the instantaneous mixture $\mathbf{x}_k = \mathbf{A}\mathbf{v}_k$. The mixture is a random $\mathbf{A} \in \mathbb{R}^{4\times 4}$ matrix, generated so that the condition number is approximately 3. The elements of the source vector $\mathbf{v}_k = [v_{k,1} \ldots, v_{k,N}]^T$ are the linear AR(1)-process

$$v_{k,n} = a_n v_{k-1,n} + u_{k,n}$$

where $u_{k,n}$ are iid. standard double exponential. The coefficients a_n determines the poles of the AR(1)-processes, and are randomly selected from a uniform pdf. in the interval $[-1/2\ ;\ 1/2]$.

The estimation of the matrices for the joint diagonalization is based on 1000 samples of the observation vector $\mathbf{x}_k$, and a distinct number of matrices is used in the joint diagonalization. The spatial correlation matrices $\mathbf{R}_{\mathbf{x},l}$ for the method in Section 4.2.1 are estimated for time shift $l = 0, 1, \ldots, L-1$. The $\mathbf{\Psi}_{\mathbf{x}}(\mathbf{t}_l)$ matrices for the method discussed in Section 4.2.2 are estimated for $\mathbf{t}_l$, $l = 1, \ldots, L$ vectors. The vectors are generated in sets of four vectors, each set is forming an orthogonal set with random rotation, and the vectors are scaled to length $1/10$. The cumulant matrices $\mathbf{Q}(\mathbf{W}_l)$ for the method in Section 4.2.3 are a subset of the complete set of parallel cumulant slice matrices.

Monte Carlo Simulation The result of a Monte Carlo simulation using the preceding simulation setup is shown in Figure 4.13. The performance of the methods is evaluated as the average and standard

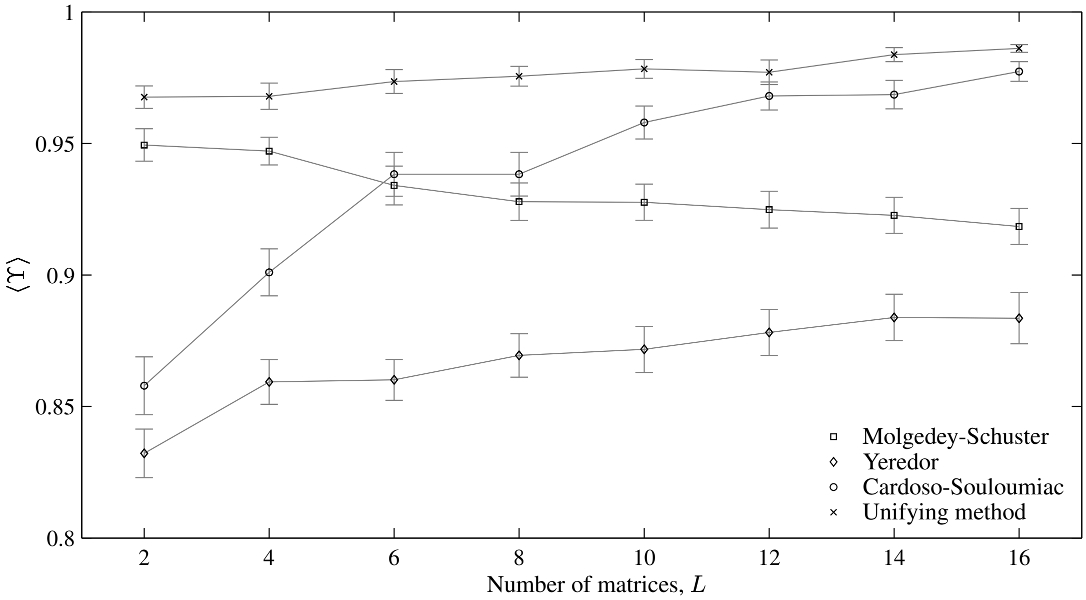


Figure 4.13 Average of the performance function in Equation 4.43 versus the number of matrices in the joint diagonalization. The errorbars indicate the standard deviation $(\sigma/10)$. Molgedey-Schuster, Yeredor, Cardoso-Souloumiac and Unifying method refer to the methods in Section 4.2.1, 4.2.2, 4.2.3 and 4.2.4 respectively.

deviation over the cost function in Equation 4.43, where the average and standard deviation are calculated over 100 Monte Carlo simulations.

The unifying method is a joint diagonalization over the set of matrices in the Molgedey-Schuster and Cardoso-Souloumiac methods. Notice that in this specific example the unifying method diagonalizes over twice as many matrices as the other methods. For all methods the matrices in the joint diagonalization are equally weighted.

For the specific simulation setup, the following interesting observations can be made. The Molgedey-Schuster method achieves a good diagonalization based on only two spatial correlation matrices. Since there is no additive noise in the observation vector and the spatial correlation matrices are estimated on a reasonable data set, the set of spatial correlation matrices are well conditioned even in the case of only two matrices. Furthermore, the sources are modelled as AR(1)-processes, and the amplitude of the resulting correlation function is therefore a constantly decaying function from lag zero. Consequently the set of spatial correlation matrices formed by a zero time shift and a one time shift is the best conditioned set of correlation matrices. Joint diagonalization over a larger set of spatial correlation matrices introduces more estimation variance in the diagonalization and the distance between the mixture and the resulting diagonalization matrix is increased. For the Yeredor and Cardoso-Souloumiac methods it is observed that the performance of the methods increases as the set of matrices in the joint diagonalization increases. For a small set of cumulant matrices the Cardoso-Souloumiac method would in general benefit from using a weighted set of cumulant slices instead of a subset of parallel cumulant slices. For the combined Molgedey-Schuster/Cardoso-Souloumiac method with a small number of matrices the joint diagonalization takes advantage of the well conditioned spatial correlation matrices, and for larger sets of matrices the method benefits from the cumulant matrices. This is a very interesting property.

Many important issues are not explored in the simulation study, among those: robustness against additive noise, non-square mixtures, influence from the eigenvalue spread of the mixture, and estimation errors in the matrices used in the joint diagonalization.

4.3 A Method Based on the Spectral Measure for Stable Distributions

In this section a novel method for independent component analysis based on the spectral measure for stable distributions is proposed. The method is based on an interpretation of the spectral measure as a probability measure for independent components as a function of the direction. Based on the spectral measure the number of independent components and the non-orthogonal bases of these components are identified. The method does not require a priori knowledge on the number of independent components, and the bases of the components can be identified even in the case of under-determined mixtures. The method has recently been published in [56] and the paper is included in Appendix A.4.

The generative model is the instantaneous mixture $\mathbf{x} = \mathbf{A}\mathbf{v}$ where $\mathbf{v}$ is the source vector, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the mixture matrix, and $\mathbf{x}$ is the observation vector. Fundamentally the blind identification of the mixture has a scale ambiguity between the nth source and the nth column of the mixture and therefore without loss of generality the columns of the mixture are normalized, $\|\mathbf{a}_{\cdot n}\|_2 = 1$ for $n = 1, \ldots, N$. The method is based on the following observation: if the source vector $\mathbf{v} = [v_1, \ldots, v_N]^T$ is an αS vector with independent $v_n \sim \alpha\mathrm{S}(\alpha, \beta_n, \gamma_n, \mu_n)$ elements for $n = 1, \ldots, N$, then the observable vector has a discrete spectral measure, and the spectral measure only has masses in the directions of the bases in the mixture. The observation follows from Proposition 3.2 and Equation 3.5. In particular, from Equation 3.5, the spectral measure for a normalized mixture is

$$\Gamma(\mathbf{s}) = \sum_{n=1}^{N} \gamma_n^{\alpha} \frac{1+\beta_n}{2} \delta(\mathbf{s} - \mathbf{a}_{\cdot n}) + \sum_{n=1}^{N} \gamma_n^{\alpha} \frac{1-\beta_n}{2} \delta(\mathbf{s} + \mathbf{a}_{\cdot n}) \tag{4.44}$$

Thus, the spectral measure reveals the bases of the mixture matrix, the number of independent components, and the scaling and skewness of each of the independent components. In general the spectral measure has masses in $2N$ directions, and for SαS sources the spectral measure is symmetric. Furthermore from Equation 4.44 it appears that there is a sign ambiguity between β_n and $\mathbf{a}_{\cdot n}$, and no ordering of the sources is provided. These are well-known inherent limitations for the blind source separation problem.

4.3.1 Estimation of the Independent Components

In practice the method is based on an estimated spectral measure, and in Section 3.3.4 a method for estimation of the spectral measure is discussed. Due to a finite number of samples of $\mathbf{x}$, observation noise and deviations from the theoretical distributions, there will be some deviations from the theoretical spectral measure in Equation 4.44. Furthermore the spectral measure is estimated in a finite number of points on the unit sphere, and this sampling of the unit sphere determines the resolution of the bases in the estimated mixture. The number of independent components is determined as the number of dominating masses in the spectral measure. This approach is severely sensitive to noise in the estimated spectral measure and to poorly scaled sources.

There are many open issues of the proposed method, among those is the sampling characteristics of the spectral measure, a reconstruction formula for the discrete spectral measure and a strategy for the selection of independent components. Even though there are many open issues, the method is still very interesting from a theoretical point of view, because it is showing that the bases of the mixture can be identified even in the case of an under-determined mixture. To the knowledge of the author this method is the first blind method for identifying the bases for continuous pdf.s in a general under-determined mixture.

The identification of the bases of the under-determined mixture leaves a new unsolved problem behind: how to reconstruct the sources. For square and over-determined mixtures the sources can be reconstructed from the inverse of the mixture, however this approach does not apply to the case of an under-determined mixture.

4.3.2 Examples with Under-determined Mixtures

The practical applicability and properties of the method are still to be examined. However, for demonstrating the method, two examples with synthetic generated signals are given next. Furthermore an example with acoustical signals and a synthetic mixture is given in Section 5.4.

Example with 2×4 **Mixture** Let $v_n \sim \mathrm{S}\alpha\mathrm{S}(1.2, 1)$ for $n = 1, \dots, 4$, and

$$\mathbf{A} = \begin{bmatrix} \cos(\theta_1) & \cos(\theta_2) & \cos(\theta_3) & \cos(\theta_4) \\ \sin(\theta_1) & \sin(\theta_2) & \sin(\theta_3) & \sin(\theta_4) \end{bmatrix}$$

with $\theta_1 = (\frac{\pi}{9})$, $\theta_2 = (\frac{2\pi}{9})$, $\theta_3 = (\frac{3\pi}{9})$, and $\theta_4 = (\frac{4\pi}{9})$. The spectral measure is estimated from 10^4 samples of $\mathbf{x}$, sampling radius is $r = 1/2$, $\mathbf{t}_n = r[\cos\left(\pi\frac{n-1}{L}\right) \quad \sin\left(\pi\frac{n-1}{L}\right)]^T$, and $L = 90$ sample points. A typical example of an estimated spectral measure is shown in Figure 4.14(a) and it appears that

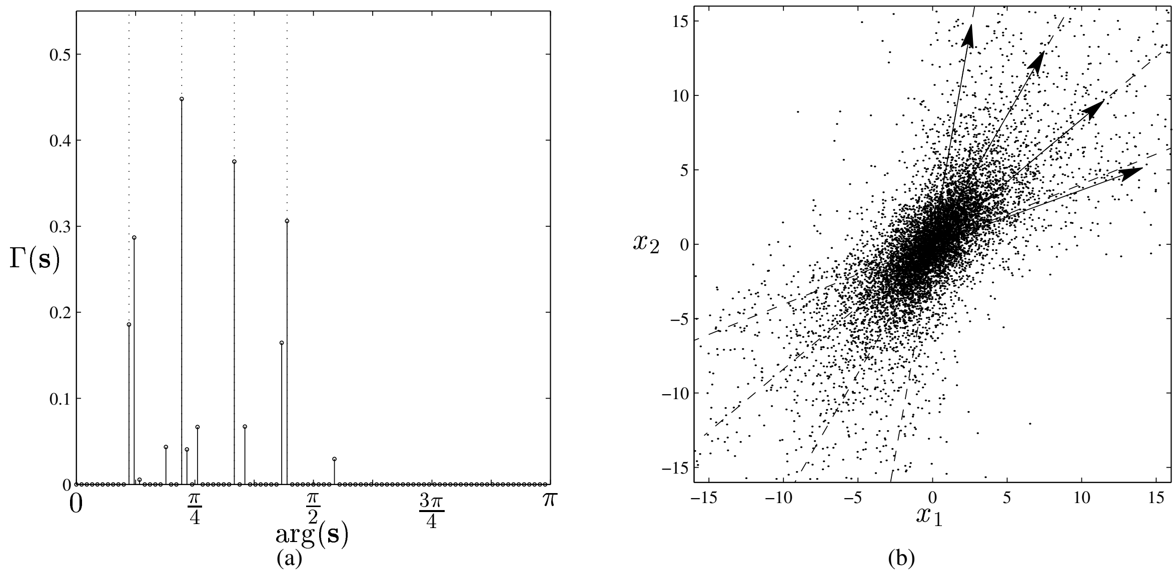


Figure 4.14 Example with 2×4 Mixture. (a) Empirical spectral measure, $\Gamma(\mathbf{s})$, for the SαS random vector $\mathbf{x}$. The estimation is based on $K = 10000$ observations of $\mathbf{x}$, sampling radius $r = \frac{1}{2}$, and $L = 90$. The directions of the basis vectors of $\mathbf{A}$ are indicated by dotted lines. (b) Scatter plot of $\mathbf{x}$. The basis vectors of $\mathbf{A}$ are depicted by vectors, and the directions in which the spectral measure has significant masses are depicted by dashed lines.

the distribution of masses for the spectral measure is very specific in the 4 directions corresponding to the directions of the basis of $\mathbf{A}$. In Figure 4.14(b) the scatter plot for the observations of $\mathbf{x}$ is depicted, and the basis vectors of the mixing matrix and the directions in which the spectral measure has dominating mass are shown. Three basis vectors are correctly identified, and the fourth basis vector is one point away from the right value.

Example with 3×10 **Mixture** Let $v_n \sim \mathrm{S}\alpha\mathrm{S}(1.2, 1)$ for $n = 1, \dots, 10$, and let the mixing matrix, $\mathbf{A}$, be a random generated $\mathbb{R}^{3\times10}$ matrix with normalized columns ($\mathbf{a}_{\cdot n}^T \mathbf{a}_{\cdot n} = 1$ for $n = 1, \dots . N$). The observable vector is thus a mixture of 10 independent SαS (1.2,1) source signals. Consider the random generated mixture matrix

$$\mathbf{A} = \begin{bmatrix} -.5459 & -.8122 & .8310 & -.3968 & .7888 & -.2928 & -.6410 & .9892 & .3424 & .6206 \\ -.3830 & -.3688 & .0415 & .6125 & -.6059 & .4526 & -.2929 & .0956 & -.2380 & -.6883 \\ .7452 & -.4521 & -.5547 & .6837 & -.1030 & -.8423 & .7095 & -.1112 & .9089 & -.3756 \end{bmatrix}$$

Since $\mathbf{A}$ has normalized columns the (m,n)th entry of $|\mathbf{A}^T\mathbf{A}|$ expresses the angle between the mth and nth column of $\mathbf{A}$. From

$$|\mathbf{A}^T\mathbf{A}| = \begin{bmatrix} \boxed{1.000} & 0.247 & 0.882 & 0.491 & 0.275 & 0.641 & \boxed{0.990} & 0.659 & 0.581 & 0.355 \\ 0.247 & \boxed{1.000} & 0.439 & 0.212 & 0.370 & 0.451 & 0.307 & 0.788 & 0.601 & 0.080 \\ 0.882 & 0.439 & \boxed{1.000} & 0.683 & 0.687 & 0.242 & 0.938 & 0.887 & 0.229 & 0.695 \\ 0.491 & 0.212 & 0.683 & \boxed{1.000} & 0.754 & 0.182 & 0.560 & 0.410 & 0.339 & 0.924 \\ 0.275 & 0.370 & 0.687 & 0.754 & \boxed{1.000} & 0.418 & 0.401 & 0.733 & 0.320 & 0.945 \\ 0.641 & 0.451 & 0.242 & 0.182 & 0.418 & \boxed{1.000} & 0.542 & 0.152 & \boxed{0.973} & 0.176 \\ \boxed{0.990} & 0.307 & 0.938 & 0.560 & 0.401 & 0.542 & \boxed{1.000} & 0.740 & 0.495 & 0.462 \\ 0.659 & 0.788 & 0.887 & 0.410 & 0.733 & 0.152 & 0.740 & \boxed{1.000} & 0.214 & 0.589 \\ 0.581 & 0.601 & 0.229 & 0.339 & 0.320 & \boxed{0.973} & 0.495 & 0.214 & \boxed{1.000} & 0.034 \\ 0.355 & 0.080 & 0.695 & 0.924 & 0.945 & 0.176 & 0.462 & 0.589 & 0.034 & \boxed{1.000} \end{bmatrix}$$

it appears that column 1 and 7 of $\mathbf{A}$ are very close, and also column 6 and 9 are, except from a sign, quite similar. Thus, we may expect some difficulty in distinguishing between these columns.

The spectral measure is estimated in a finite number of points on the unit sphere, and the sampling grid is constructed as follows. Take basis in an isocahedron - an isocahedron is a polyhedron with 20 equilateral triangles. Each of the equilateral triangles is divided into K^2 equilateral triangles with side length $1/K$ of the original triangles. The sampling grid on the unit sphere is then obtained as the projection of the resulting vertices onto the unit sphere.

The spectral measure is estimated in 642 points on the unit sphere (corresponding to $K = 4$), the estimation is based on 10^5 samples of $\mathbf{x}$, and using Equation 3.28 the sampling radius is $r = (N\gamma)^{-1/\alpha} = 0.1468$. The following estimated mixture matrix is obtained

$$\begin{array}{cc} \hat{\mathbf{A}} = & \begin{bmatrix} .8550 & .8507 & .3618 & .2629 & -.5254 & -.3618 & .7534 & -.9867 & -.6163 & -.6457 \\ .3312 & -.0000 & -.5878 & -.3892 & -.3817 & .2629 & -.6440 & -.0823 & .6893 & -.3128 \\ .3991 & -.5257 & -.7236 & .8829 & .7604 & -.8944 & -.1330 & .1401 & .3809 & .6966 \end{bmatrix} \\ \text{index} & \begin{array}{cccccccccc} (2) & (3) & (4) & (6) & (1) & (9) & (5) & (8) & (10) & (7) \end{array} \end{array}$$

Considering $|\hat{\mathbf{A}}^T\mathbf{A}|$ the estimated mixture can be compared to the true mixture. From

$$|\hat{\mathbf{A}}^T\mathbf{A}| = \begin{bmatrix} 0.296 & \boxed{0.997} & 0.502 & 0.136 & 0.432 & 0.436 & 0.361 & 0.833 & 0.576 & 0.152 \\ 0.856 & 0.453 & \boxed{0.998} & 0.697 & 0.725 & 0.193 & 0.918 & 0.899 & 0.186 & 0.725 \\ 0.511 & 0.250 & 0.677 & \boxed{0.998} & 0.716 & 0.237 & 0.573 & 0.382 & 0.393 & 0.900 \\ 0.663 & 0.469 & 0.287 & 0.260 & 0.352 & \boxed{0.996} & 0.571 & 0.124 & 0.985 & 0.099 \\ \boxed{0.999} & 0.223 & 0.874 & 0.494 & 0.261 & 0.659 & 0.988 & 0.640 & 0.602 & 0.348 \\ 0.569 & 0.601 & 0.206 & 0.306 & 0.352 & 0.978 & 0.479 & 0.233 & \boxed{0.999} & 0.069 \\ 0.263 & 0.314 & 0.673 & 0.784 & \boxed{0.998} & 0.400 & 0.388 & 0.698 & 0.290 & 0.960 \\ 0.674 & 0.768 & 0.901 & 0.436 & 0.742 & 0.133 & 0.755 & \boxed{0.999} & 0.191 & 0.608 \\ 0.356 & 0.074 & 0.694 & 0.927 & 0.943 & 0.171 & 0.463 & 0.586 & 0.028 & \boxed{1.000} \\ 0.991 & 0.324 & 0.936 & 0.540 & 0.391 & 0.539 & \boxed{0.999} & 0.746 & 0.486 & 0.447 \end{bmatrix}$$

the ordering of the columns between $\hat{\mathbf{A}}$ and $\mathbf{A}$ is obtained. The index beneath the estimated mixture corresponds to the column index of the true mixture matrix. It appears that the 10 bases of the 3-dimensional observation vector have been identified.

4.4 Summary

A probabilistic generative model is established for instantaneous mixtures of temporal independent sources, for instantaneous mixtures of sources with temporal dependence, and for convolutive mixtures of temporal independent sources. The maximum likelihood method for blind signal separation is derived for all three generative models.

The likelihood functions for SαS and NIG source distributions are examined, and it is shown that the likelihood function has maximum when the estimation matrix equals the inverse of the mixture matrix even for relatively large deviations between the model distribution and the true distribution of the sources. The non-linear estimation function is depicted for SαS and NIG distributions, and it is shown that the stochastic gradient method is useful for blind signal separation for heavy tail distributed source signals.

Several methods for independent component analysis based on joint diagonalization of estimated matrices are considered. A new method based on spatial covariation matrices that are suitable for SαS signals is proposed, and the theoretical basis for using the method based on spatial covariance matrices in the case of αS signals is established. A unifying method, that combines any methods that can be formulated as a joint diagonalization problem, is proposed. A simulation example compares the methods and it is demonstrated that the unifying method in general has superior performance compared to the other methods.

A novel method for independent component analysis based on the discrete spectral measure for αS distributions is proposed. The method can indentify the number of independent component and the non-orthogonal bases of the components. The method does not require a priori knowledge on the number of independent components, and the bases of the components can be identified even in the case of under-determined mixtures. Simulations show that the method can identify the bases of under-determined mixtures.

CHAPTER 5

APPLICATIONS IN AUDIO SIGNAL PROCESSING

This chapter will focus on application in audio signal processing of the methods developed in the previous chapters. The chapter is organized as follows.

In section 5.1 modelling of audio signals using the SαS and the NIG pdf.s are discussed, and it is demonstrated that the class of SαS pdf.s are attractive for modelling a broad class of audio signals. Section 5.2 proposes a novel stochastic gradient adaptive filter that is suitable for non-Gaussian heavy tail signals. The adaptive filter is based on a signal dependent non-linear function; and it is demonstrated, using both synthetic generated signals and audio signals, that the method has superior performance compared to state-of-the-art robust adaptive filtering methods. The applicability to audio signal processing is demonstrated in an acoustic echo canceling setup.

In Section 5.3 and 5.4 the identification of the bases of instantaneous mixtures of audio signals are considered. Section 5.3 compares the methods based on joint diagonalization of matrices discussed in Section 4.2; and in Section 5.4 examples of identifying the bases of under-determined instantaneous mixtures of audio signals are given.

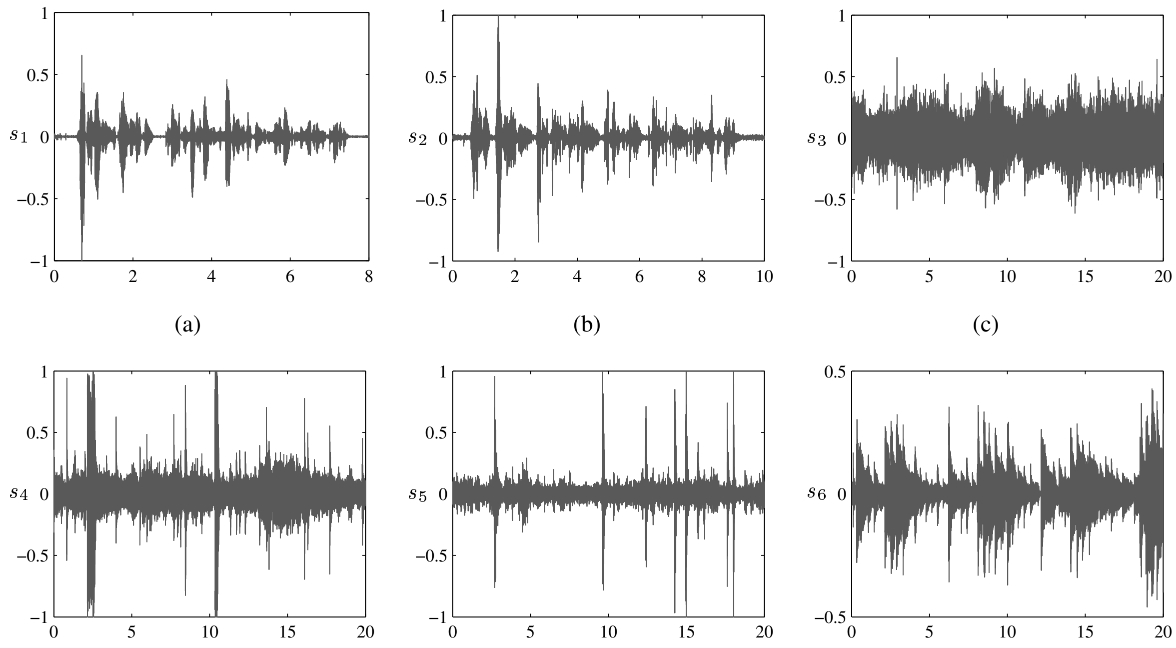
Section 5.5 demonstrate that the maximum likelihood method for blind signal separation of convolutive mixtures applies to audio signals. The observable signals are synthetic convolutive mixtures of speech signals. In Section 5.6 the maximum likelihood method for blind signal separation of convolutive mixtures is applied to microphone signals from an experimental teleconference system. It is demonstrated that two speech signals to some extent can be separated, and a spatial interpretation of the separation system is provided.

5.1 Modelling of Audio Signals Using Heavy Tail Distributions

Recently in [38] SαS distributions have been applied for modelling of audio signals with emphasis on noise signals. In [55] SαS distributions are applied to a broader class of audio signals, and it is demonstrated that the SαS density functions are suitable for modelling the short as well as the long term amplitude distributions of the audio signals.

Audio signals are in general not stationary and modelling of the amplitude pdf.s thus is strongly dependent on the time scale. In [38] it is reported that for audio signals the characteristic exponent in the SαS model is typical in the range α =1.5 – 1.6 (time window 0.7 – 15 sec.). For shorter time windows the characteristic exponent tends to have a larger value. The audio signals, thus, appear to have a less impulsive nature on a small time scale but more impulsive on a longer time scale. In [55], where a much broader class of audio signals is considered, α is found to be in the interval 0.9 – 1.8 on a long time scale (8 – 20 sec.). Furthermore the fluctuations of α on a small time scale (100 ms.) is examined. For details see the examples in [55] included in Appendix A.2.

In this section the suitability of the SαS and NIG models for modelling audio signals is discussed and the discussion is restricted to the long time scale. In Figure 5.1 six different audio signals that will be used in the following discussion are shown. Estimates of the SαS and NIG distribution parameters



(a) (b) (c) (d) (e) (f)

Signal	Description
s_1	Speech signal, male, low background noise. Sampling freq. 32 kHz., 8 sec.
s_2	Speech signal, male, low background noise. Sampling freq. 32 kHz., 10 sec.
s_3	Cocktail party background noise. Sampling freq. 44.1 kHz., 20 sec.
s_4	Background noise recorded in a kitchen. Sampling freq. 20 kHz., 20 sec.
s_5	Background noise recorded in an office. Sampling freq. 44.1 kHz., 20 sec.
s_6	Music, classical guitar. Sampling freq. 44.1 kHz., 20 sec.

(g)

Figure 5.1 Audio signals. Plot (a) – (f) is the time sequence of the audio signals s_1 – s_6, the abscissa is time [sec.], the ordinate is the amplitude normalized with the dynamical range of the AD-converter. Additional information for the audio signals is listed in the table.

for the six audio signals in Figure 5.1 are obtained using the parameter estimation methods from Section 3.3.3. An estimate of the standard deviation is obtained with the standard maximum likelihood estimator. The empirical amplitude pdf. for the six audio signals and the corresponding SαS , NIG and Gaussian probability density function fit are shown in Figure 5.2 and 5.3.

In Figure 5.2(a) – (f) the pdf.s are depicted on a linear scale, and from the figure it appears that the two heavy tail distributions has a better fit to the empirical pdf.s than the Gaussian distribution. Since the Gaussian pdf. only has a scale parameter while the SαS and the symmetric NIG pdf has both a scale and a "shape" parameter, this conclusion is reasonable. The pdf.s of the cocktail party noise signal in Figure 5.2(c) appears to be the "most" Gaussian signal. It also appears from the figure that the SαS distribution has the best fit to the amplitude histogram in all cases except for the case with cocktail party noise.

To examine the model's capability of modelling the tails of the signals, Figure 5.3(a) – (f) shows

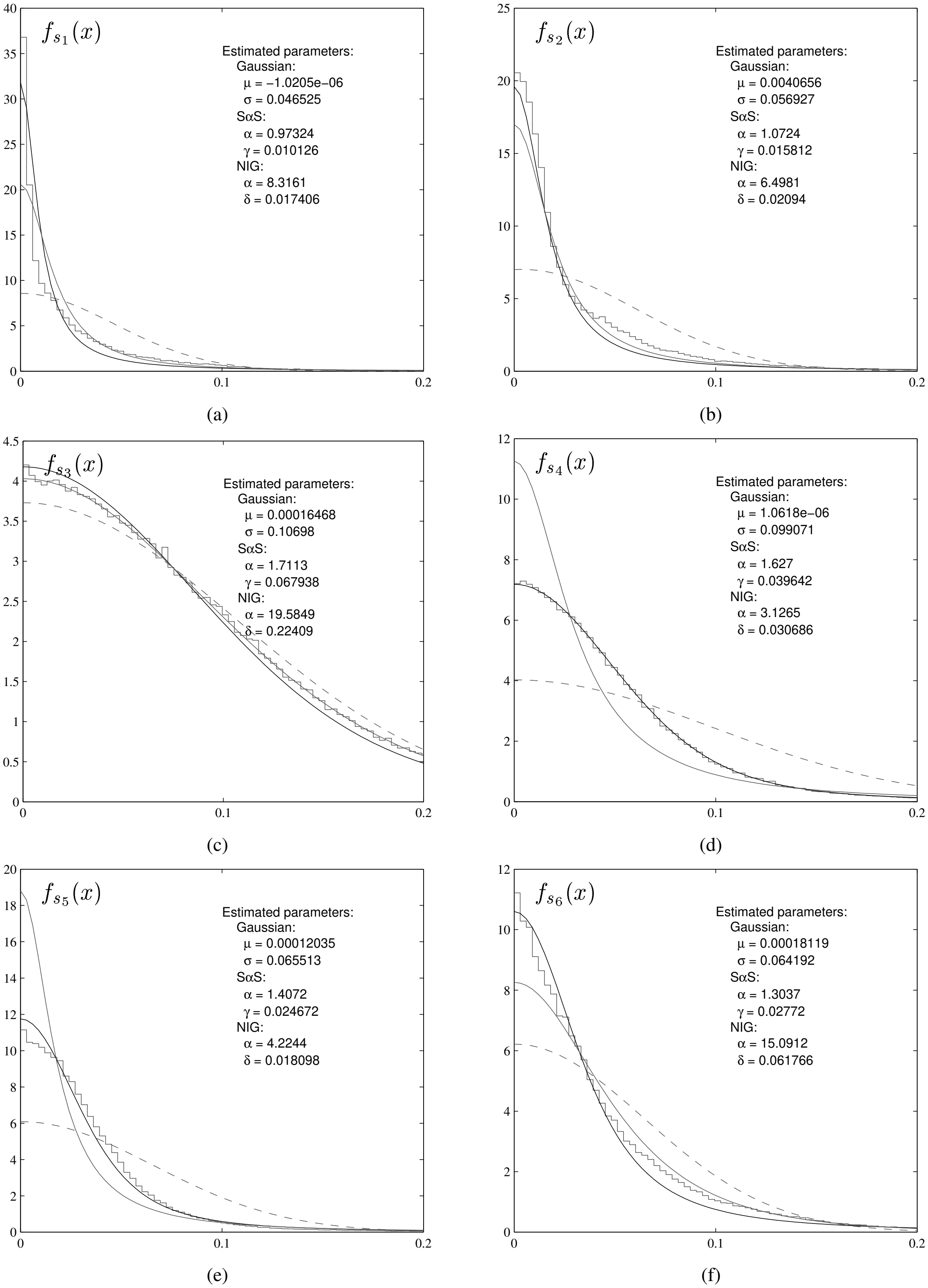


Figure 5.2 Empirical pdf.s for audio signals and the corresponding SαS (solid), NIG (gray) and Gaussian (dashed) pdf.s fits. (a) Speech signal. (b) Speech signal. (c) Cocktail party background noise. (d) Background noise from a kitchen (e) Background noise from an office. (f) Classical guitar.

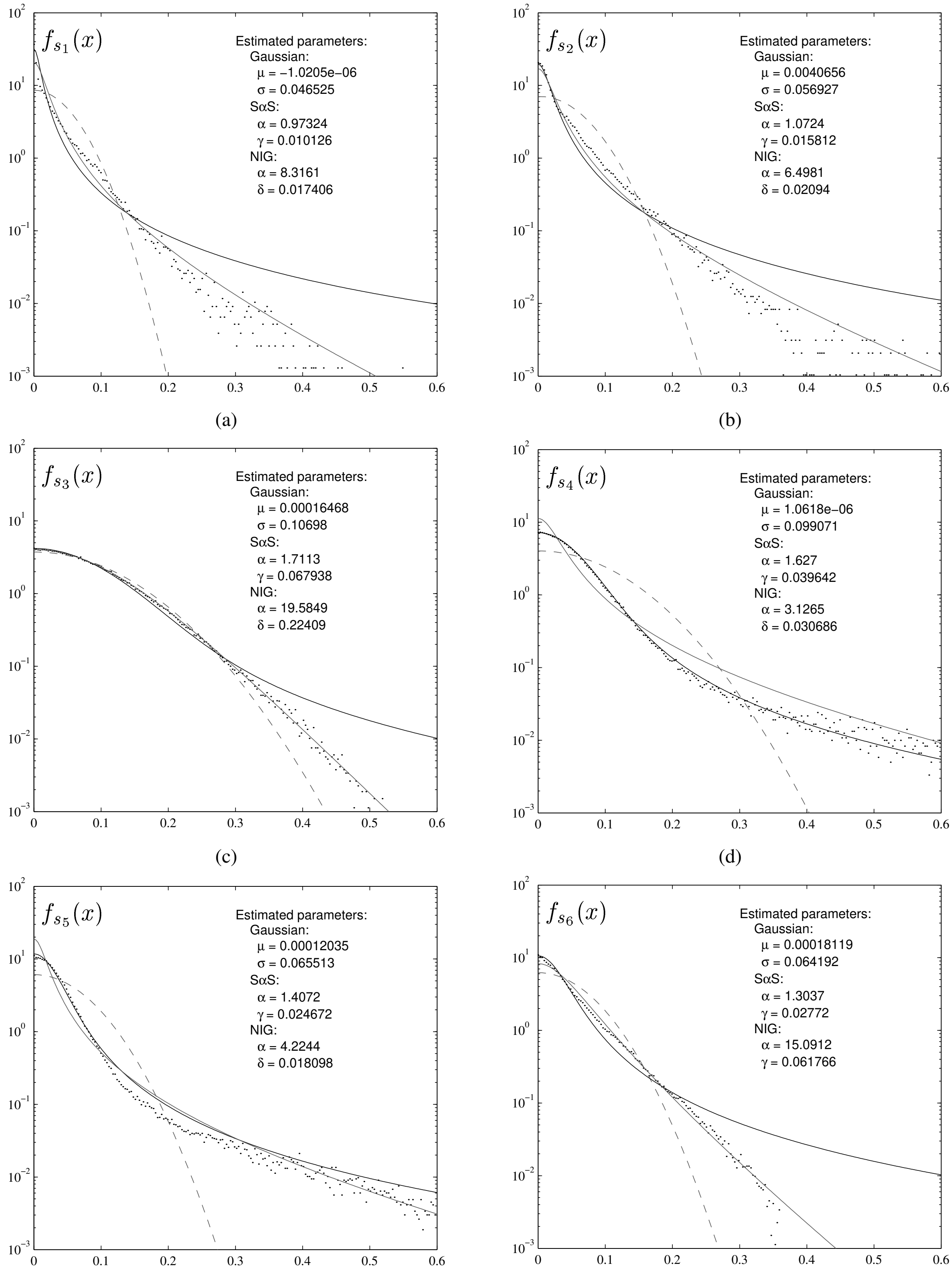


Figure 5.3 Empirical pdf.s for audio signals and the corresponding SαS (solid), NIG (gray) and Gaussian (dashed) pdf.s fits. Notice that the ordinate is logarithmic. (a) Speech signal. (b) Speech signal. (c) Cocktail party background noise. (d) Background noise from a kitchen (e) Background noise from an office. (f) Classical guitar.

the pdf.s depicted on a logarithmic scale. From the figure it appears that the tail fit of the NIG model is better than the SαS model for the two speech signals, the cocktail party background noise and the classical music signal. The tail fit of the SαS model is better than the NIG model for the kitchen and office background noise signals.

Because the SαS model has the best fit to the amplitude histogram where the signals has most probability mass, and because both the SαS and NIG model has reasonable tail fits, it is believed that the SαS model in general is the best suited model for modelling audio signals.

5.2 A Robust Adaptive Filtering Method for Non-Gaussian Signals

In this section a stochastic gradient adaptive filtering method that is suitable for non-Gaussian heavy tail signals is proposed. The update scheme of the adaptive filter is based on the non-parametric amplitude transformation proposed in Section 2.2.2. The method is published in [54] and the paper is included in Appendix A.1.

Adaptive Filtering of Heavy Tail Signals Consider an adaptive FIR filter[1] with input u_n. The estimation error, at time n, of the filter is

$$e_n = d_n - \mathbf{w}^T \mathbf{u}_n \tag{5.1}$$

where $\mathbf{u}_n$ is the input signal vector, $\mathbf{w}$ is the filter coefficient vector, and d_n is the desired signal. The adaptation of the adaptive filter is based on an objective function, J, and classically the objective is to minimize the second order moment of the estimation error. The class of stochastic gradient adaptive filters has update equations on the form

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu g(e_n \mathbf{u}_n) \tag{5.2}$$

where μ is an adaptation constant, and $g(\cdot)$ is an estimate of the gradient $\mathrm{E}\left(\partial J/\partial \mathbf{w}\right)$ based on the estimation error and the input signal.

For heavy tail signals in general an objective function based on a second order moment is not suitable. Therefore a number of different robust stochastic gradient algorithms, based on a fractional lower order moments (FLOM), has been proposed [7, 12, 13, 31]. One of these is the Symmetric Least Mean P-norm (SLMP), that uses the symmetric norm

$$g(e\mathbf{u}) = (e\mathbf{u})^{\langle p/2 \rangle}$$

as an estimate of the gradient, [13]. Another subclass of stochastic gradient adaptive filters is proposed in [31], with the filter update equation

$$\left[g_q(\mathbf{u}_n)\right]_i = \frac{|u_{i,n}|^{q-1}\mathrm{sign}(u_{i,n})}{\sum_{m=1}^{L} |u_{i,n}|^q} \quad \text{for } 1 \leq q < \infty \tag{5.3}$$

where $\left[g_q(\mathbf{u})\right]_i$ denotes the ith element of the vector valued function $g_q(\mathbf{u})$. This update, for $\mu = 1$ and for a valid q, corresponds to minimizing $\|\mathbf{w}_{n+1} - \mathbf{w}_n\|_p$ subject to $d_n - \mathbf{w}_{n+1}^T \mathbf{u}_n = 0$. The relation between the q-norm of the input signal vector, and the p-norm of the coefficient vector update, $\|\mathbf{w_{n+1}} - \mathbf{w_n}\|_p$, is $1/p + 1/q = 1$, [31]. Thus, the update in Equation 5.3 provides the minimum p-norm of the coefficient

[1] See Haykin's classical textbook [47] for a thorough introduction to adaptive filter theory.

vector update, and for $q = 2$ the algorithm is the classical NLMS algorithm. Finally, [7] proposed a generalization to the NLMP with the update equation

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{e_n^{\langle a \rangle}}{\|\mathbf{u}_n\|_{qa}^{qa} + \lambda} \mathbf{u}_n^{\langle (q-1)a \rangle} \tag{5.4}$$

where, in the presence of a SαS process, $0 < a \leq \alpha - 1$ and $1 \leq q$. The update is motivated by observing that it corresponds to a gradient descent adaptation approach to the objective function $J = \mathrm{E}\left(|e_n|^{a+1}\right)$. The update in Equation 5.4 reduces to the NLMP update if a and q are chosen as $p - 1$ and p/a. The update in equation 5.3 is a special case of Equation 5.4 for $a = 1$ and $\lambda = 0$

The stochastic gradient adaptive filter proposed here is conceptually different from these methods. The idea is to base the update of the filter on a non-linear amplitude transformation of the signals that ensures the existence of the second order moment. From this idea the update equation on the form of Equation 5.2 is

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu g_{\mathbf{u}}(\mathbf{u}_n)\, g_e(e_n) \tag{5.5}$$

where $g_{\mathbf{u}}$ and g_e are non-linear transformations that transforms arbitrary pdf.s into a Gaussian pdf. The convergence characteristic of the method, using the median orthogonality criterion proposed in [12], is discussed in [54]. The structure of the adaptive filter is depicted in Figure 5.4.

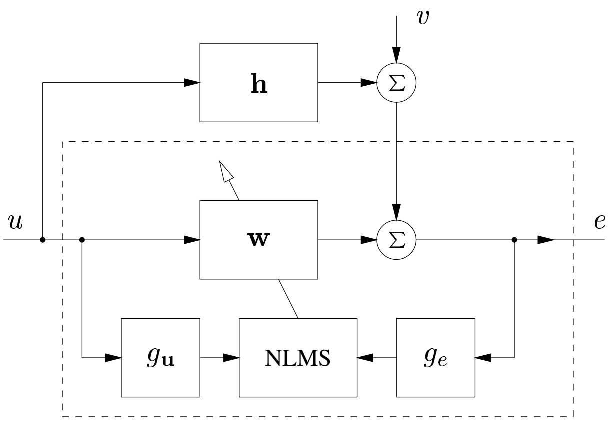


Figure 5.4 Adaptive filter based on non-linear transformations.

The non-linear transformations are based on the empirical transformation to a uniform pdf., a transform with the inverse on the Gaussian distribution function, and a scaling. The scaling is necessary since all the scaling information is lost in the empirical density transformation; for details see [54]. The number of observations N used in the empirical transformation determines the quantization level of the signals, e.g. if the transformation is based on 1024 observations then the quantization corresponds to a 10-bit AD-converter. Furthermore, the coherent nature of the signal must be taken into consideration when choosing the number of observations that the transformation is based on.

Simulation with Synthetic Signals The setup used for evaluation is depicted in Fig. 5.4, where $\mathbf{h}$ is a 10th order low pass filter. The input signal u and the observation noise signal v is generated as iid. SαS random variables. The proposed algorithm is compared to three other algorithms, representing state of the art in robust adaptive algorithms. The update equations and the adaptation constants used in the simulation study are listed in Table 5.1. These constants have considerable influence on the characteristics

Algorithm		Update Equations	Adaptation constants
Aydíns method	[7]	$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{\mathbf{u}^{\langle (q-1)a \rangle}}{\|\mathbf{u}\|_{qa}^{qa} + \lambda} e^{\langle a \rangle}$	$\mu = .01,\ p = q = 1.1,\ a = \alpha_v - 1.1$
NSLMP	[13]	$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{1}{\|\mathbf{u}\|_p^p + a} (e\mathbf{u})^{\langle p/2 \rangle}$	$\mu = .01,\ p = .9\alpha_v,\ a = 10^{-6}$
NLMP	[13]	$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{1}{\|\mathbf{u}\|_p^p + a} e^{\langle p-1 \rangle} \mathbf{u}$	$\mu = .01,\ p = .9\alpha_v,\ a = 10^{-6}$
Proposed meth.	[54]	$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu g_{\mathbf{u}}(\mathbf{u}_n)\, g_e(e_n)$	$\mu = .05,\ N = 120$

Table 5.1 Update equations and adaptation constants used in the simulation study with SαS signals.

of the algorithms, and the constants are chosen in such a way that all the algorithms are robust and stable in the wide range of signals under consideration in this simulation study. Notice, that in the advantage of the algorithms used for comparison, the norms used in these algorithms are adjusted in accordance with the characteristic exponent of the applied signals. The performance of the algorithms is evaluated as the modelling error

$$10 \log_{10} \mathrm{E}\left\{(\mathbf{w} - \mathbf{h})^T (\mathbf{w} - \mathbf{h}) / (\mathbf{h}^T \mathbf{h})\right\}$$

Consider the case where the source signal u and the observation noise signal v are iid. Cauchy distributed signals. The convergence of the modelling error is depicted in Fig. 5.5(a). The NLMP, NSLMP and Aydíns method converges to about the same misadjustment level, but the proposed method has a lower misadjustment. Aydíns method has a considerable slower convergence than the other algorithms.

For the case where the source signal u is an iid. Gaussian signal, and the observation noise signal v is an iid. Cauchy distributed signal, the convergence of the modelling error is depicted in Fig. 5.5(b). Two important changes are observed: the misadjustment level is about 20 dB higher and the convergence speed is slower. The γ-ratio is not a suitable measure between variables of different distributions, so even though the γ-ratio between the signals is still equal to one, the empirical SNR is very poor. In Fig. 5.5(c) the modelling error versus signal to interference dispersion ratio is depicted. The dispersion of the source signal varies in the interval $\gamma_u^2 = [0.1, 100]$, and for the interfering signal the dispersion is constant, $\gamma_v = 1$. For low dispersion ratios the performance of the algorithms are much the same, but for increasing dispersion ratio the algorithms has increasing performance. For large dispersion ratios the proposed method has considerable lower modelling error than the other algorithms.

In Fig. 5.5(d) the modelling error versus the characteristic exponent is depicted. The methods used for comparison are almost independent of α, since in this simulation study the norm is adjusted in accordance with the applied signal. Even though the proposed method does not require any a priori knowledge of the input signals distribution, it out-performs all the algorithms used for comparison. This is a remarkable robustness characteristic.

Acoustical Echo Canceling of Speech Signals Consider the described adaptive algorithms used in an acoustical echo canceling setup. The setup is as depicted in Fig. 5.4 where v is the local speaker and u is the remote speaker. The acoustical path, modeled by the filter $\mathbf{h}$, is the same 10th order filter as in the previous simulation study. The adaptation constants used in this example are listed in Table 5.2. For the NSLMP, NLMP and Aydíns method the p-norm and μ are set to relative small values in order to obtain stability. The signals are speech signals sampled at 8 kHz.. The input signals and the modelling error result of the simulation is shown in Fig. 5.6. The simulation example contains several difficulties: the signals are non-stationary and contains temporal correlation, the amplitude density is non-Gaussian, and

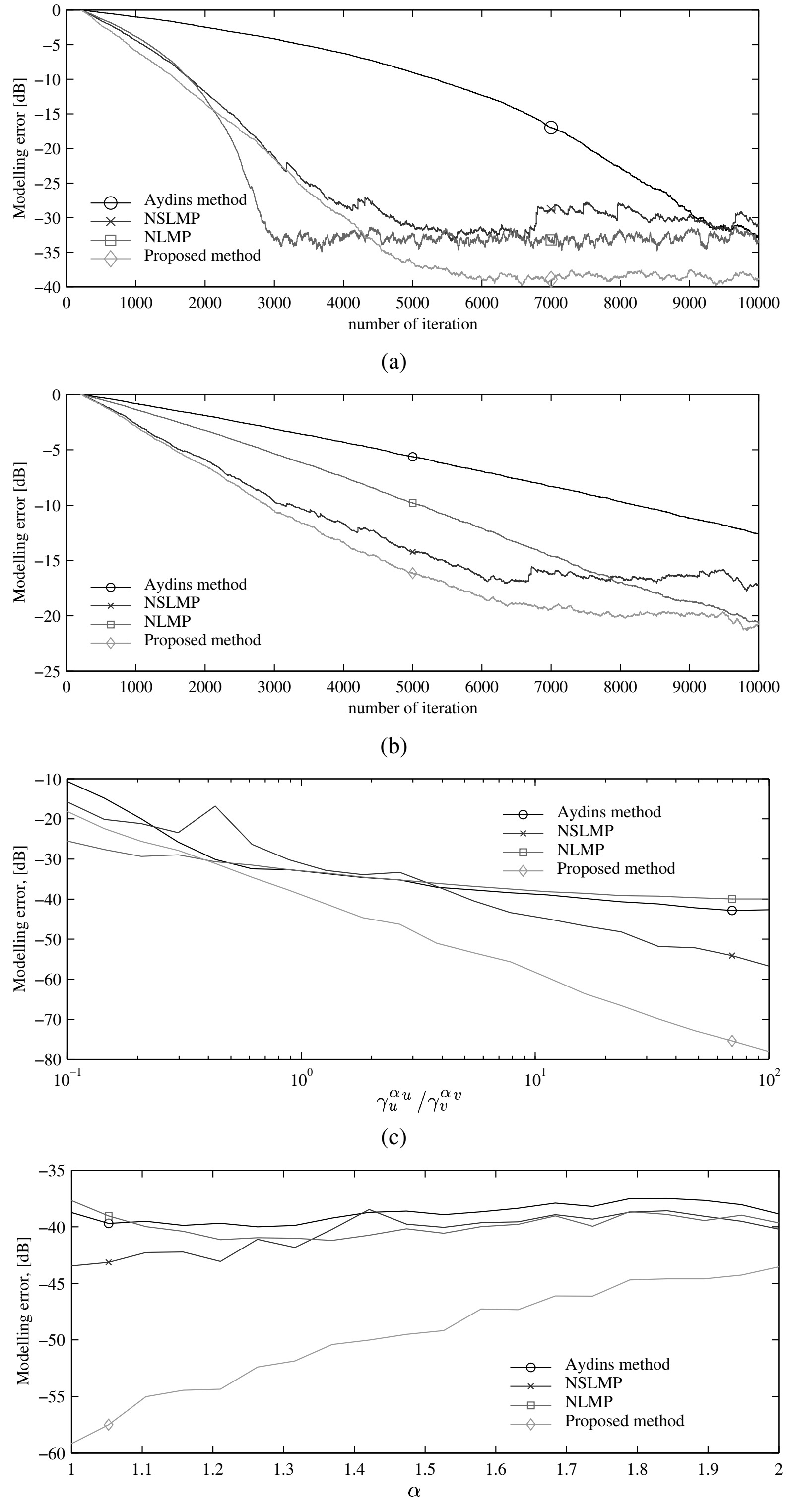


Figure 5.5 Simulation results. Mean over 10 Monte Carlo runs. (a) Modelling error for $\alpha_u = \alpha_v = 1$ and $\gamma_u = \gamma_v = 1$. (b) Modelling error for $\alpha_u = 2$, $\alpha_v = 1$ and $\gamma_u = \gamma_v = 1$. (c) Modelling error versus signal to interference dispersion ratio, $\gamma_u^{\alpha_u} / \gamma_v^{\alpha_v}$. Characteristic exponent $\alpha_u = \alpha_v = 1$. (d) Modelling error versus the characteristic exponent, α. The dispersion is $\gamma_u^{\alpha} = 1$ and $\gamma_v^{\alpha} = 0.1$.

Algorithm	Adaptation constants
Aydíns method	$\mu = .01, p = q = 1.1, a = .1$
NSLMP	$\mu = .01, p = 1.44, a = 10^{-6}$
NLMP	$\mu = .01, p = 1.44, a = 10^{-6}$
Proposed method	$\mu = .1\ , N = 1000$

Table 5.2 Adaptation constants used in the acoustical echo cancellation for speech signals

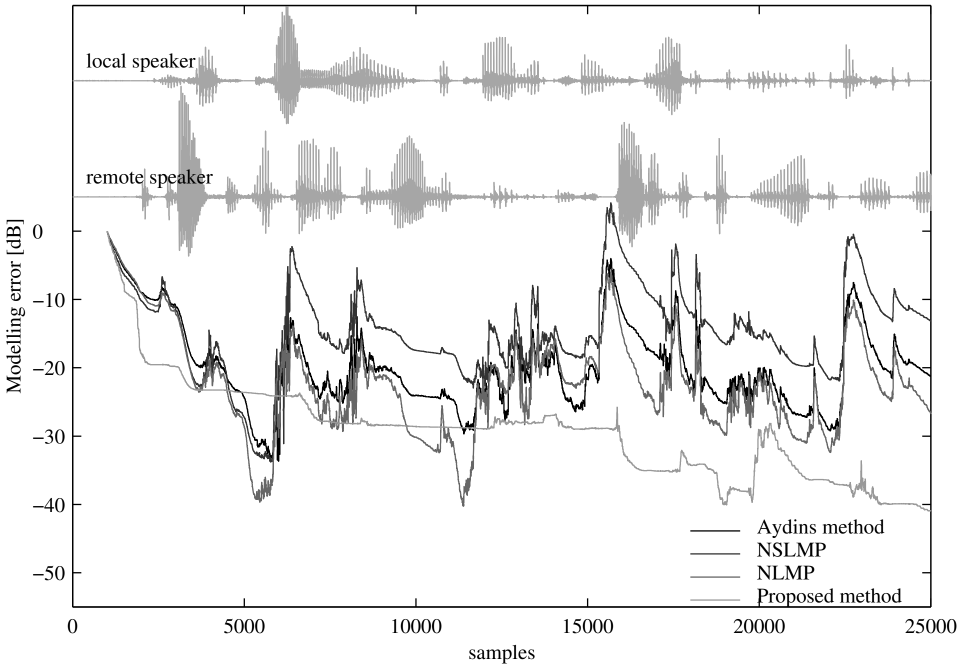


Figure 5.6 Modelling error, in acoustical echo-canceler in cross-talk environment, for the 4 methods.

the dispersion ratio between the signals are strongly varying. It appears, from Fig. 5.6, that the proposed method has slower convergence but apart from that, it has considerable better performance compared with the other methods. In the intervals where the local speaker is active, which corresponds to a poor SNR, the algorithms, except for the proposed method, lose the tracking of the filter. For these methods a control strategy is necessary. The proposed method has considerable better performance in the case of double-talk situations and environments with high background noise levels.

5.3 Separation of Instantaneous Mixture of Speech Signals

Blind separation of signals modelled as instantaneous mixtures are interesting in a number of practical application. The instantaneous mixture model is not a realistic model for blind separation of audio signals from an microphone array, however, it is of general interest to examine the performance of the methods from Chapter 4 for real world signals. In this and the following section the separation of audio signals, where the mixture model is an instantaneous mixture, is examined.

Here the methods based on joint diagonalization from Section 4.2 is examined by means of a simulation example with an instantaneous mixture of audio signals with additive observation noise. The mixture model is $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$, where $\mathbf{x}$ is the observable vector, $\mathbf{s}$ is the source vector, $\mathbf{A} = [\cos(\theta_1) \quad \cos(\theta_2); \sin(\theta_1) \quad \sin(\theta_2)]$ is the mixture matrix with $\theta_1 = \pi/16$ and $\theta_2 = \pi/4$, and $\mathbf{n}$ is the noise vector. The source vector contains 64000 samples (time interval 0 – 2 sec.) of the two speech signals s_1 and s_2 from Figure 5.1. The noise signals are synthetical generated as AR(1)-processes with iid. double exponential process excitations and with pole-location $z = 0.4$. The amplitude of the noise signal is scaled so that the resulting SNR is 6 dB.

In Figure 5.7 the result of the simulation is shown for diagonalization over 4 different sets of matrices. In the scatter plots of the observations the bases of the mixture and the identified bases are depicted. The vectors indicate the bases of the mixture, and the dashed lines indicate the bases of the solution to the joint diagonalization. In Figure 5.7(a) the bases identified by an ordinary Molgedey-Schuster method

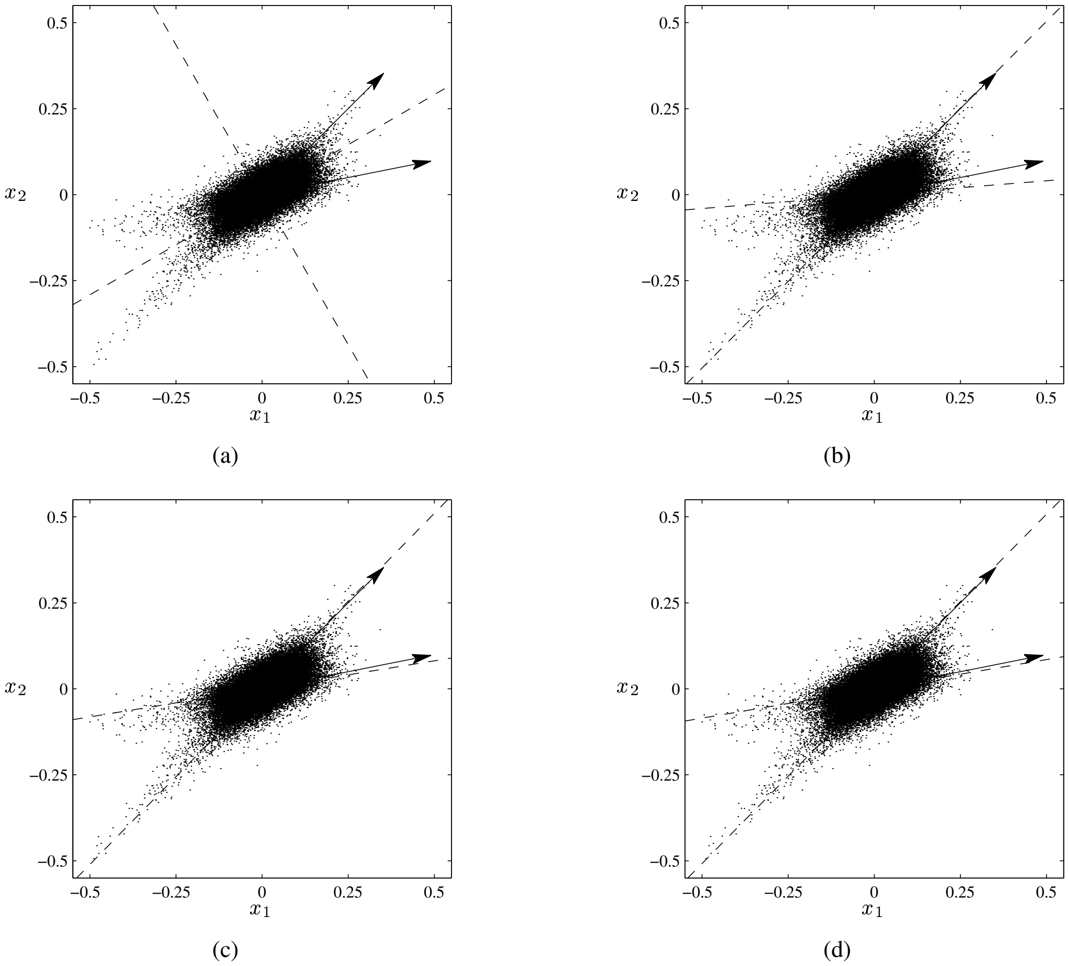


Figure 5.7 Scatter plots for instantaneous mixture of audio signals with additive noise. The vectors indicate the bases of the mixture, and the dashed lines indicate the bases of the solution to the joint diagonalization. (a) Ordinary Molgedey-Schuster. (b) Molgedey-Schuster, joint diagonalization over 4 spatial covariance matrices. (c) Cardoso-Souloumiac, joint diagonalization over 4 cumulant matrices. (d) Combined method, joint diagonalization over 2 spatial covariance matrices and 2 cumulant matrices.

are depicted. The bases are identified as the solution to the eigenvalue decomposition of the spatial covariance matrix with lag 1 divided by the spatial covariation matrix with lag 0. The observation noise biases the diagonal of the spatial covariance matrices, and from Figure 5.7(a) it is seen that the directions of the identified bases are very different from the direction of the bases of the mixture matrix. In Figure

5.7(b) the bases obtained as the joint diagonalization over 4 spatial covariance matrices with lag 0, 2, 4 and 6 are depicted. Because the correlation of the speech signals is stronger than the correlation of the observation noise, the Molgedey-Schuster benefits from diagonalizing over a larger set of matrices with a larger time-window. In Figure 5.7(c) the bases obtained as the joint diagonalization over the matrices of the 4 parallel cumulant slices are depicted. It appears that the identified bases of the observable signals are very close to the bases of the mixture matrix, and for this particular simulation setup the Cardoso-Souloumiac method has better performance than the Molgedey-Schuster method. In Figure 5.7(d) the bases obtained as a joint diagonalization over a combination of spatial covariance matrices and cumulant matrices are depicted. The diagonalization is determined over a set of 4 matrices; the set consist of 2 spatial covariance matrices with lag 2 and 4, and 2 matrices of parallel cumulant slices (the (2,1)th and (1,1)th parallel cumulant slice). It appears that the identified bases are very close to the bases of the mixture matrix, and the combination of spatial covariance matrices and cumulant matrices yield the best performance. This observation is consistent with the conclusion from Section 4.2.5.

5.4 Under-determined Mixture of Audio Signals

In this section two examples of the identification of the bases of an under-determined synthetic mixture of audio signals are given using the method based on the spectral measure described in Section 4.3. The first example is taken from [56] included in Appendix A.4, and the second example uses three of the audio signals examined in Section 5.1.

The audio signals in the first example are three speech signals with sampling frequency 8 kHz. The audio signals are mixed into two observable signals with the mixture matrix $\mathbf{A} = [1 \ 1/\sqrt{2} \ 1/\sqrt{2}; 0 \ 1/\sqrt{2} \ -1/\sqrt{2}]$. The spectral measure is estimated in $L = 90$ points based on $N = 23000$ samples, sampling radius $r = 50$, and characteristic exponent $\alpha = 1.2$. In Figure 5.8(a) the estimated spectral measure is depicted; the directions of the bases of $\mathbf{A}$ are indicated by dotted lines, and it appears that the distribution of masses is very specific in the three directions corresponding to the directions of the bases in $\mathbf{A}$. In Figure 5.8(b) the scatter plot of the observation signal is depicted; the bases of $\mathbf{A}$ are indicated by vectors, and the directions in which the spectral measure has the largest masses are indicated by dashed lines. In this example the bases of the under-determined mixture is completely revealed.

The second example represents a more typical example of the performance that can be expected when using the method based on the spectral measure on acoustical signals. The source audio signals are two speech signals and a background noise signal (s_1, s_2 and s_4 from Figure 5.1), and the mixture matrix is $\mathbf{A} = [1.4142 \ \ 1.9616 \ \ 0.1951; 1.4142 \ \ 0.3902 \ \ 0.9808]$. The spectral measure is estimated in $L = 90$ points based on $N = 64000$ samples, sampling radius $r = 15.33$, and characteristic exponent $\alpha = 1.14$. The characteristic exponent is determined using the parameter estimation method described in Section 3.3.3, and the sampling radius is determined using Equation 3.28 where γ and α are determined using the parameter estimation method from Section 3.3.3.

In Figure 5.9(a) the estimated spectral measure is depicted. Compared to the spectral measure in Figure 5.8(a) the spectral measure in Figure 5.9(a) is more ”noisy” and the directions in which the spectral measure has masses are not as specific and close to the directions of the bases of the mixture. In Figure 5.9(b) the scatter plot of the observation signal is depicted; the bases of $\mathbf{A}$ are indicated by vectors, and the directions in which the spectral measure has the largest masses are indicated by dashed lines.

5.5 Separation of a Synthetic Convolutive Mixture of Speech Signals

In this section it is demonstrated that the maximum likelihood method for blind separation of convolutive mixtures, described in Section 4.1.4, can be used for separation of mixtures of audio signals. The observable signals are a synthetically generated convolutive mixture of audio signals.

The source signals are the two speech signals, s_1 and s_2, depicted in Figure 5.1(a) and 5.1(b). The speech signals are decimated with a factor 4 in order to decrease the computational requirements, and the separating system is estimated over 32000 samples (time-interval 1 – 5 sec. of the speech signals). The adaptation constant is $\mu = 10^{-3}$ and the update is using a simple line search where the adaptation coefficient is multiplied with 2 for every line search. The convolutive filters have filter length $L = 20$, and the coefficients of the convolutive filter are generated as iid. Gaussian random variables with

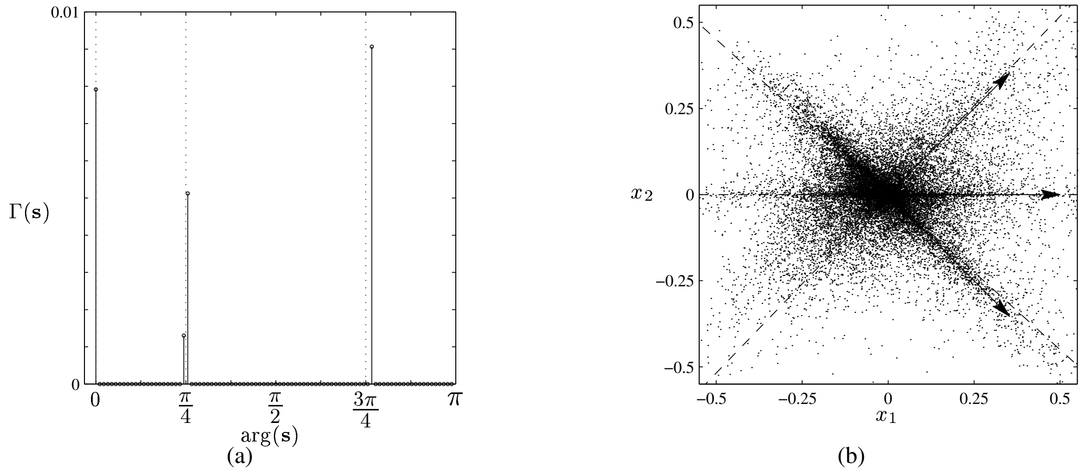


Figure 5.8 (a) Spectral measure for the observable signal vector. The dotted lines indicate the direction of the bases in the mixture matrix, $\mathbf{A}$. (b) Scatter plot, x_2 versus x_2. The bases of $\mathbf{A}$ are indicated by vectors, and the directions in which the spectral measure has the largest masses are indicated by dashed lines.

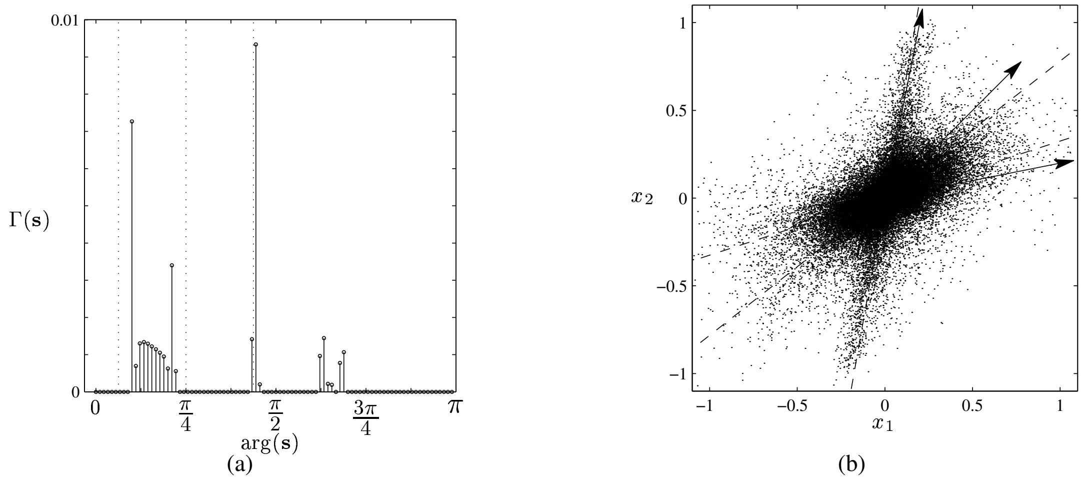


Figure 5.9 (a) Spectral measure for the observable signal vector. The dotted lines indicate the direction of the bases in the mixture matrix, $\mathbf{A}$. (b) Scatter plot, x_2 versus x_2. The bases of $\mathbf{A}$ are indicated by vectors, and the directions in which the spectral measure has the largest masses are indicated by dashed lines.

exponentially decreasing variance. In Figure 5.10 the impulse responses of the convolutive filters for this specific simulation example are depicted.

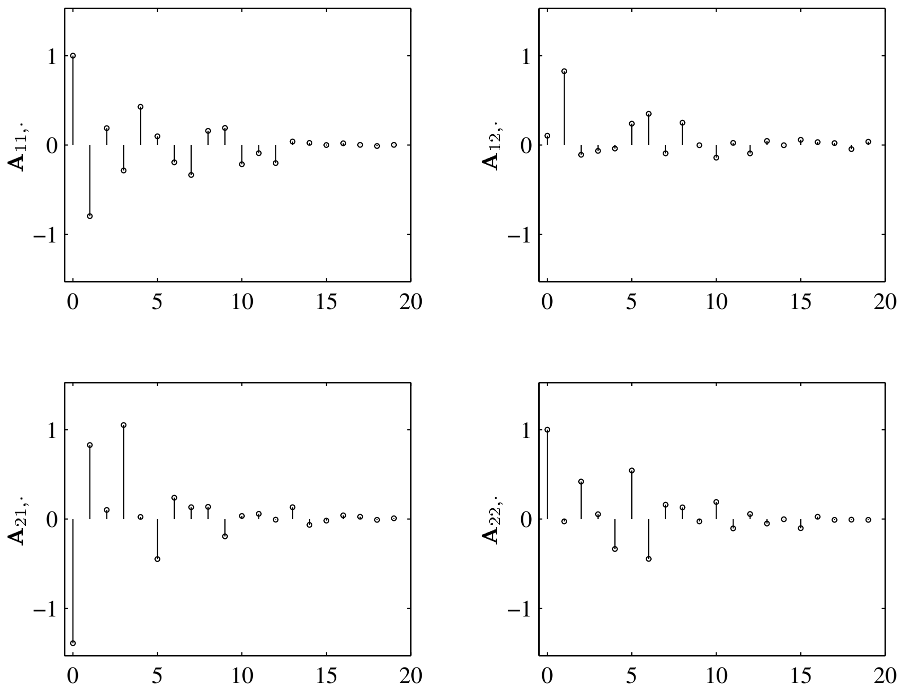


Figure 5.10 Impulse responses for the convolutive filters. $\mathbf{A}_{mn,\cdot}$ denote the coefficients in the filter from input s_n to observation x_m.

The impulse responses of the separating system are depicted in Figure 5.11; and the complete impulse responses from the source signals to the estimated signals, i.e. the mixing filters convolved with the separating filters, are depicted in Figure 5.12. From Figure 5.12 it appears that the complete impulse responses have a large gain in the anti-diagonal and a small gain in the diagonal, thus, signal separation has been obtained. However, the blind signal separation for convolutive mixture has a convolutive ambiguity, and from Figure 5.12 it appears that the complete impulse response is not close to an impulse in the anti-diagonal. This is because the correlation structure from the source signals is partly adapted in the deconvolution filters. This convolutive ambiguity was not observed in the simulation example with iid. in Section 4.1.4 because the source signals were without temporal correlation.

It is useful to make a frequency domain interpretation of the complete transfer function, and in Figure 5.13 the amplitude of the transfer function is depicted in the frequency domain. It appears that the complete transfer function has a relatively large gain in the anti-diagonal and a relative low gain in the diagonal, but the transfer functions in the diagonal are not constant over frequency. Thus, the convolutive ambiguity corresponds to a frequency weighting of the separated signals.

In the upper right plot in Figure 5.13 the amplitude of the frequency response from source number 2 to output channel number 1 is depicted with a solid line, and the amplitude of the frequency response from source number 1 to output channel number 1 is depicted with a dashed line. It appears that source signal number 1 is attenuated with approximately 10 dB compared to source signal number 2. A similar conclusion is obtained for channel 2 in the lower left plot.

It might be instructive to consider the scatter plots of the signals, and in Figure 5.14(a), Figure 5.14(b), and Figure 5.14(c) the scatter plots of the source signals, the observable signals, and the estimated signals respectively are depicted.

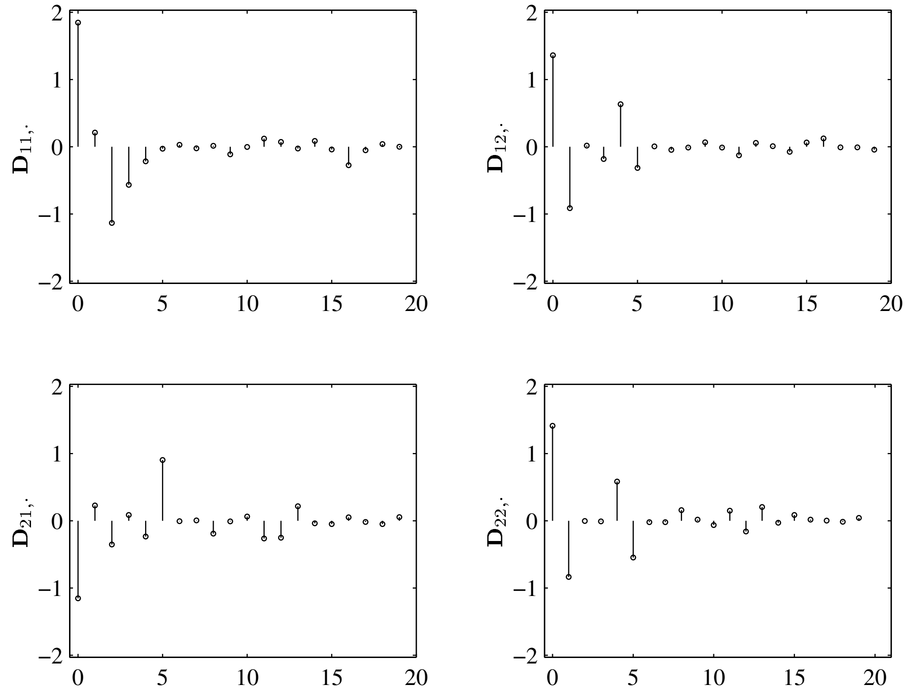


Figure 5.11 Impulse responses for the deconvolution filters. $\mathbf{D}_{mn,\cdot}$ denote the coefficients in the filter from observation x_n to output u_m.

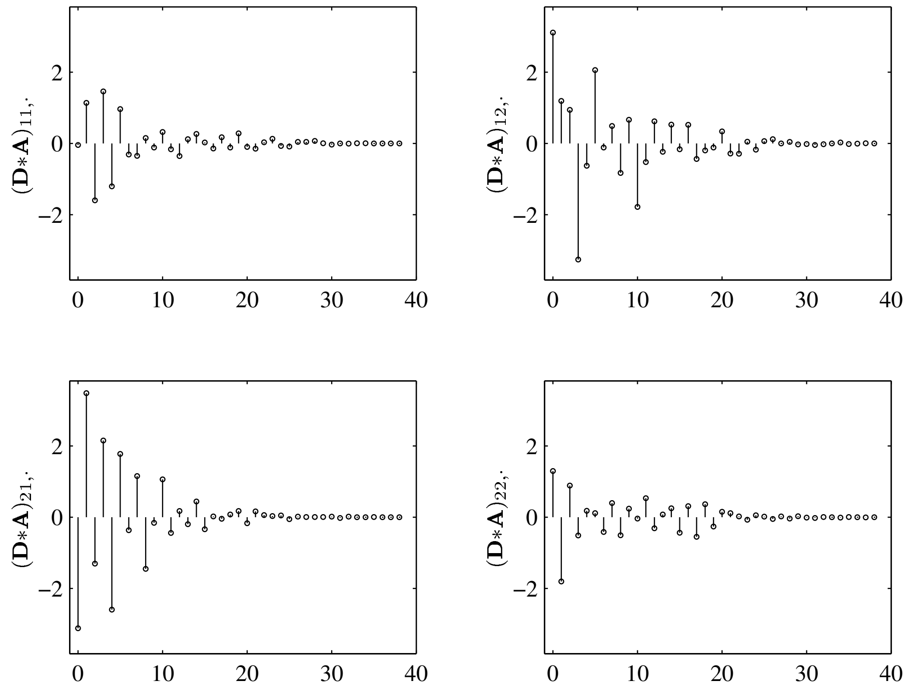


Figure 5.12 Impulse responses for the mixing filters convolved with the separating filters. $(\mathbf{D} * \mathbf{A})_{mn,\cdot}$ denote the overall impulse response from input s_n to output u_n.

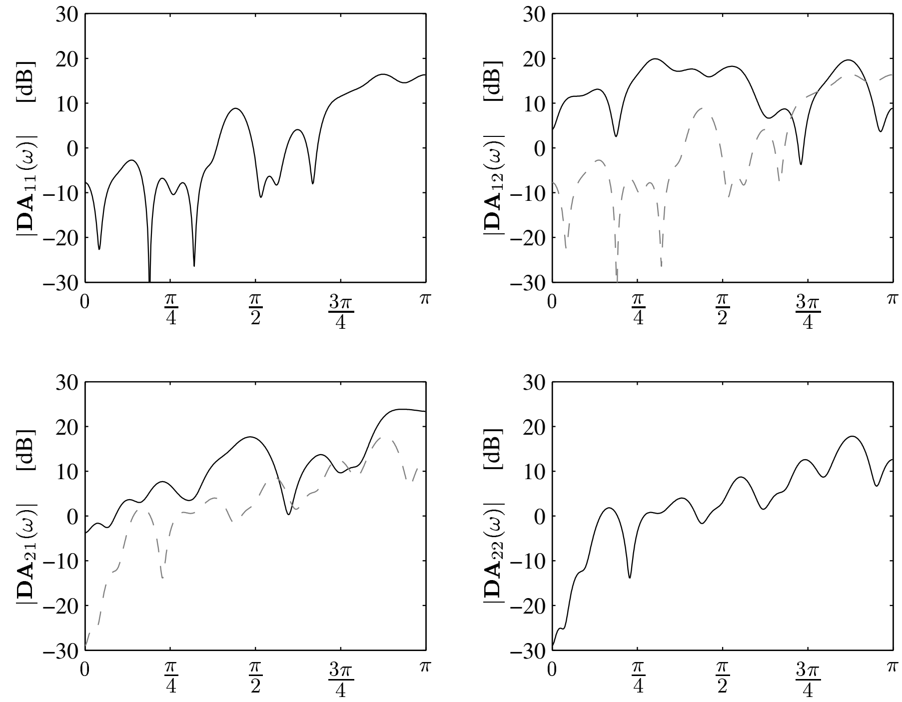


Figure 5.13 Amplitude of the frequency responses for the complete transfer function. $\mathbf{DA}_{mn}(\omega)$ denote the frequency response from input s_n to output u_n. The dashed lines in the upper right and lower left plot depict the frequency response of $|\mathbf{DA}_{11}(\omega)|$ and $|\mathbf{DA}_{22}(\omega)|$ respectively.

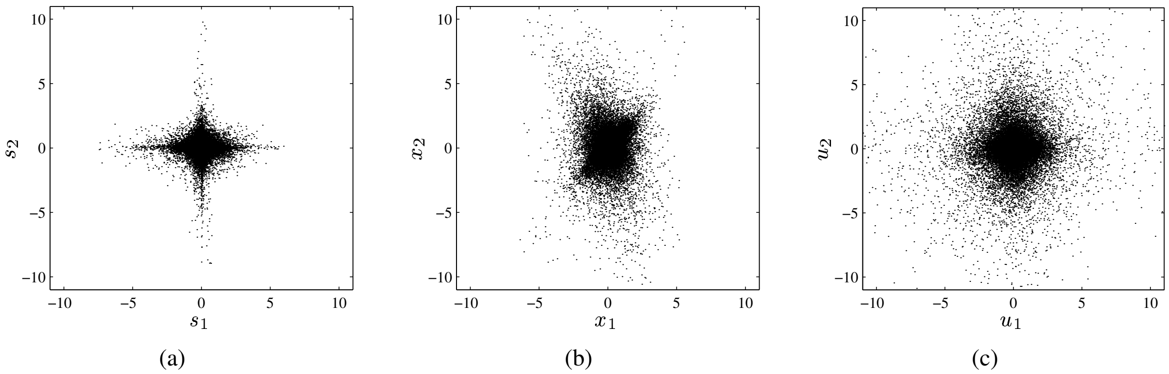


Figure 5.14 Scatter plots (a) Source signals, s_2 versus s_1. (b) Observation signals, x_2 versus x_1. (c) Estimated signals, u_2 versus u_1.

The main drawback of the time-domain maximum likelihood blind signal separation method is that it has slow convergence. In this example the estimation is stopped after 500 iterations, and in Figure 5.15 the empirical log-likelihood function is depicted. Faster convergence is reported in [6] for the frequency-domain and hybrid methods.

Plots of the source signals, the observable signals, and the estimated signals are given in Figure 5.16. From a visual inspection of the signals in Figure 5.16 the source signals can be recognized in the estimated source signals, however, a simple listening experiment is a much stronger evaluation of the signal separation than the graphical evaluation. From a listening experiment it appears that the

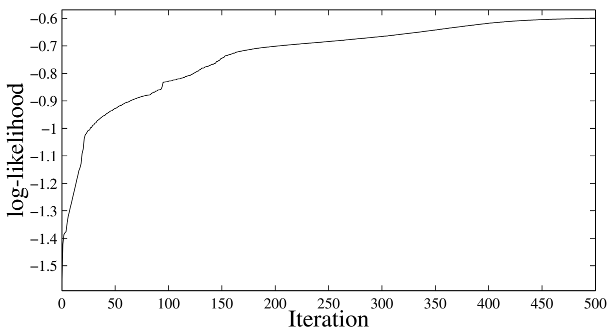


Figure 5.15 Empirical log-likelihood function for the blind signal sepration of a convolutive mixture of speech signals.

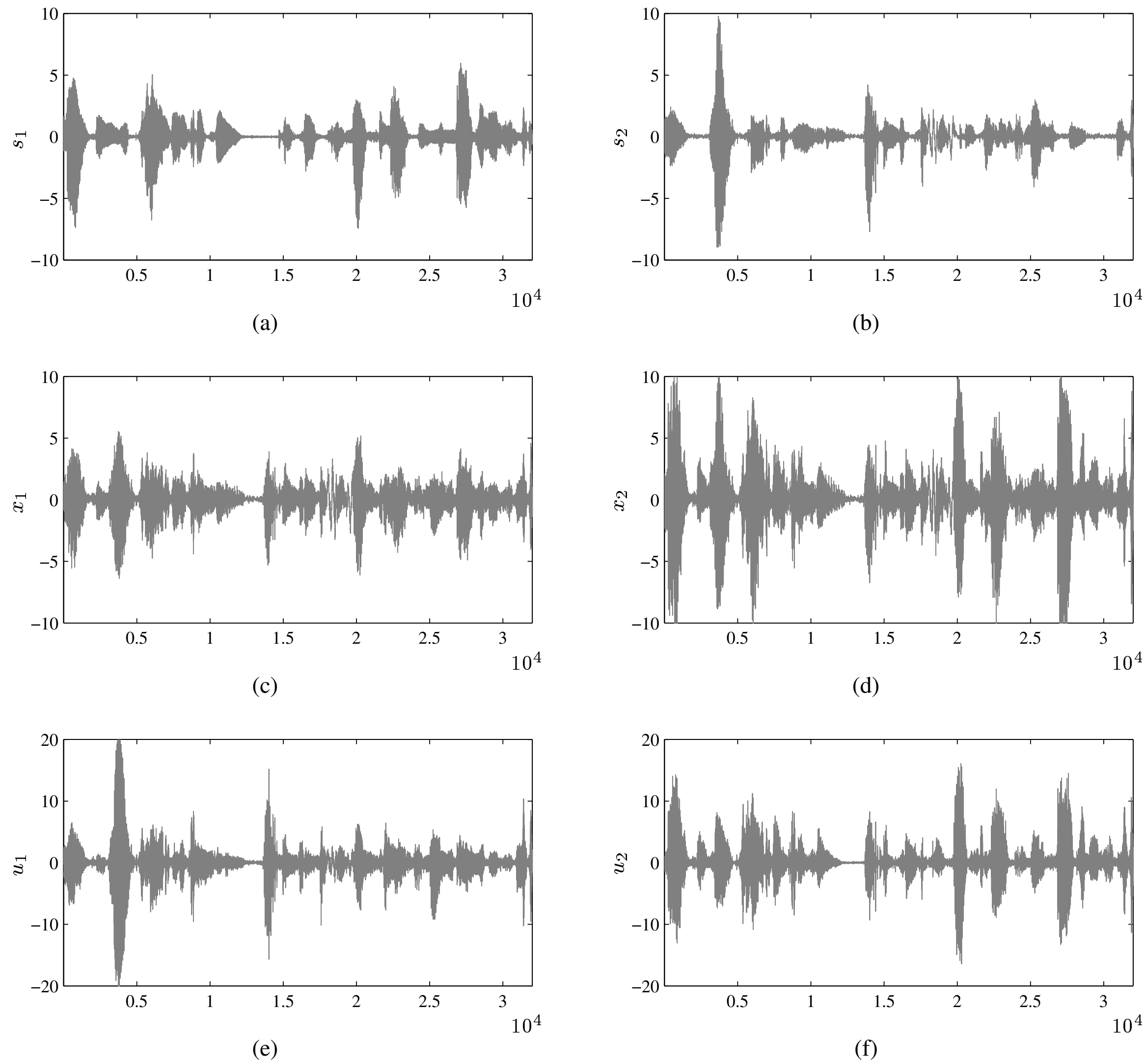


Figure 5.16 (a)–(b) Source signals. (c) Observable signal in channel 1. (d) Observable signal in channel 2. (e)–(f) Estimated output signals.

speech signals have not been completely separated, however, each of the speech signals appears to be considerably stronger in one channel than in the other.

5.6 Separation of a Real World Mixture of Speech Signals

In this section the maximum likelihood method for blind separation of convolutive mixtures, described in Section 4.1.4, is applied to signals from an experimental teleconference system[2]. The signals are sampled from 2 microphones with 135 mm distance and with sampling frequency 32 kHz.. Two speakers are located in the room approximately 150 cm from the microphone array, and both speakers are active simultaneously.

The microphone signals are decimated with a factor 4, and the deconvolution system is estimated based on 16000 samples corresponding to a time-window of 2 sec.. The adaptation constant is $\mu = 10^{-3}$ and the update is using a simple line search where the adaptation coefficient is multiplied with 2 for every line search. The convergence of the likelihood function is slow, and therefore the estimation is stopped after 1000 iterations. The impulse response of the estimated deconvolution system is depicted in Figure 5.17. Since the mixing system is unknown the complete impulse response cannot be determined.

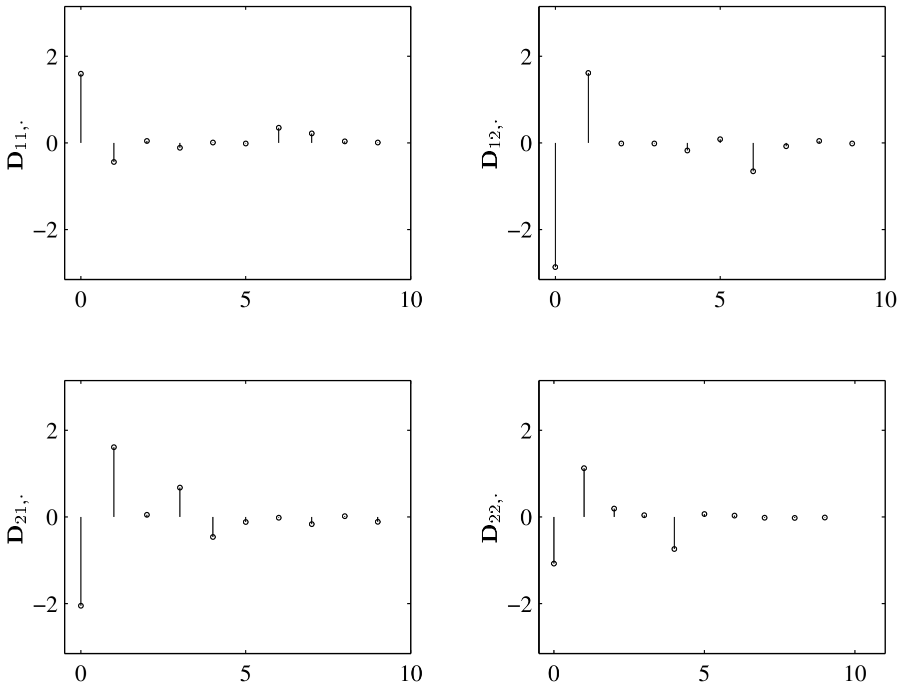


Figure 5.17 Impulse responses for the deconvolution filters. $\mathbf{D}_{mn,\cdot}$ denote the coefficients in the filter from observation x_n to output u_m.

The microphone array structure is known and therefore the spatial responses of the deconvolution filters can be determined. The spatial response for the mth channel of the deconvolution system for a linear microphone array is given as the spatio-temporal Fourier transformation

$$G_m(\psi, \omega) = \sum_{n=1}^{N} \sum_{l=0}^{L-1} \mathbf{D}_{mn,l} \exp\left(-j\omega l \frac{1}{F_s}\right) \exp\left(-j2\pi(n-1)\sin(\psi)\frac{d}{\lambda}\right)$$

[2]The experimental teleconference system is depicted in Figure 1.3.

where ψ is the azimuth angle relative to the plane orthogonal to the microphone-array [rad], ω is the frequency [rad/sec], F_s is the sampling frequency [sec^{-1}], d is the microphone distance [m], and λ is the wave length [m], see [57, 50]. The spatial response of the separation system provides an intuitive interpretation of each channel of the separating system; i.e. sources in directions where the amplitude of the spatial response is small are attenuated, while sources in directions where the amplitude of the spatial response is large are present in the channel. In an anechoic environment signals are completely separated if the separating system is able to place a zero in the amplitude of the spatial response in the direction of all but one of the source signals. The maximal number of zeros in the spatial response is $2(N-1)$ where N is the number of microphones in the array. By extending the number of microphones in the array the separation system has the freedom to place more zeros in the spatial response. Thus, in general better signal separation can be obtained when using more than 2 microphones.

In Figure 5.18 the amplitude of the spatial response, $|G_m(\psi,\omega)|$, of the 2 channels of the estimated separation system is depicted at frequency $\omega = 2\pi 10^3$. The speakers are located approximately in the directions 0° and -60° respectively. Figure 5.18(a) shows the spatial response for channel 1, and it appears that the separation system has low gain in the direction of the first speaker (0°) and large gain in the direction of the second speaker (-60°). Figure 5.18(b) shows the spatial response for channel 2, and it appears that the gain is large in the direction of the first speaker (0°) and low in the direction of the second speaker (-60°). From this spatial interpretation of the separating system it is expected that a reasonable signal separation is obtained.

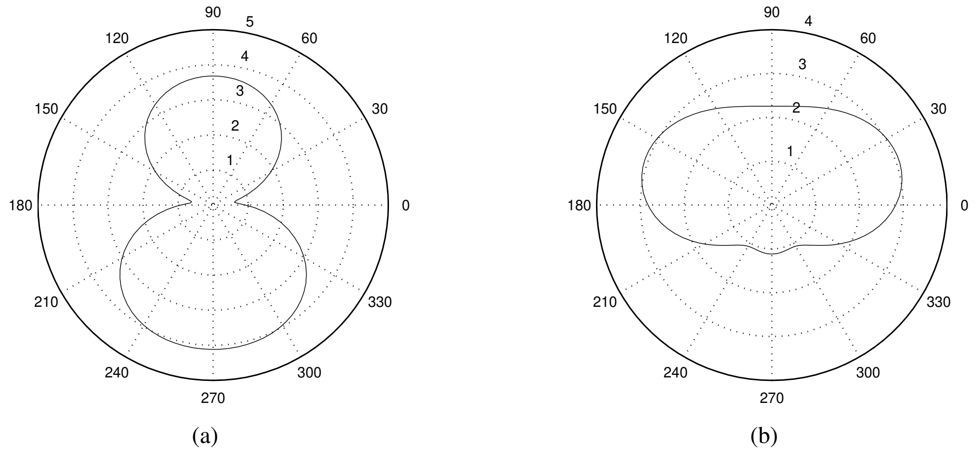


Figure 5.18 Plot of the amplitude of the spatial response, $|G_m(\psi,\omega)|$ for $\omega = 2\pi f$ and $f = 1$ kHz., in polar coordinates. (a) Channel 1, $|G_1(\psi,\omega)|$. (b) Channel 2, $|G_2(\psi,\omega)|$.

In Figure 5.19(a) the scatter plot of the observation signals are depicted. In Figure 5.19(b) the scatter plot of the estimated output signals are depicted, and the estimated signals seems to be independent. In Figure 5.20 the observable and estimated signals are depicted, and on this large time-scale the observable signals are almost indistinguishable. It is not possible to evaluate the signal separation from the plots of the estimated signals; one simple way to evaluate the signal separation is simply to listen to the estimated signals. A simple listening experiment shows that the two speech signals to some extent have been separated, but both speakers are still present in both of the estimated signals.

The general experience from the work with signals from the experimental teleconference system is that the signals can be separated to some extent. In the scenario where two speakers are speaking simultaneously it is possible to isolate the individual speakers and a significant speech intelligibility is obtained. From the simple listening experiments it is experienced that the frequency weighting of the estimated signals in general was not a problem.

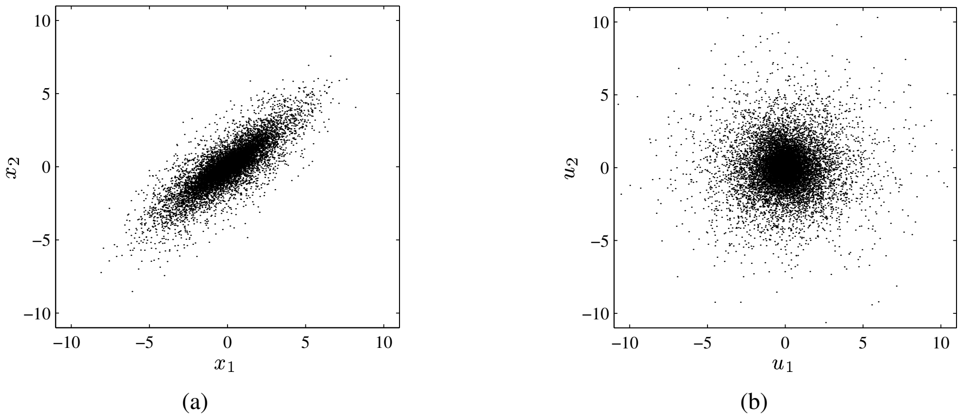


Figure 5.19 (a) Scatter plot of observation signal, x_2 versus x_1. (b) Scatter plot of output signal, u_2 versus u_1.

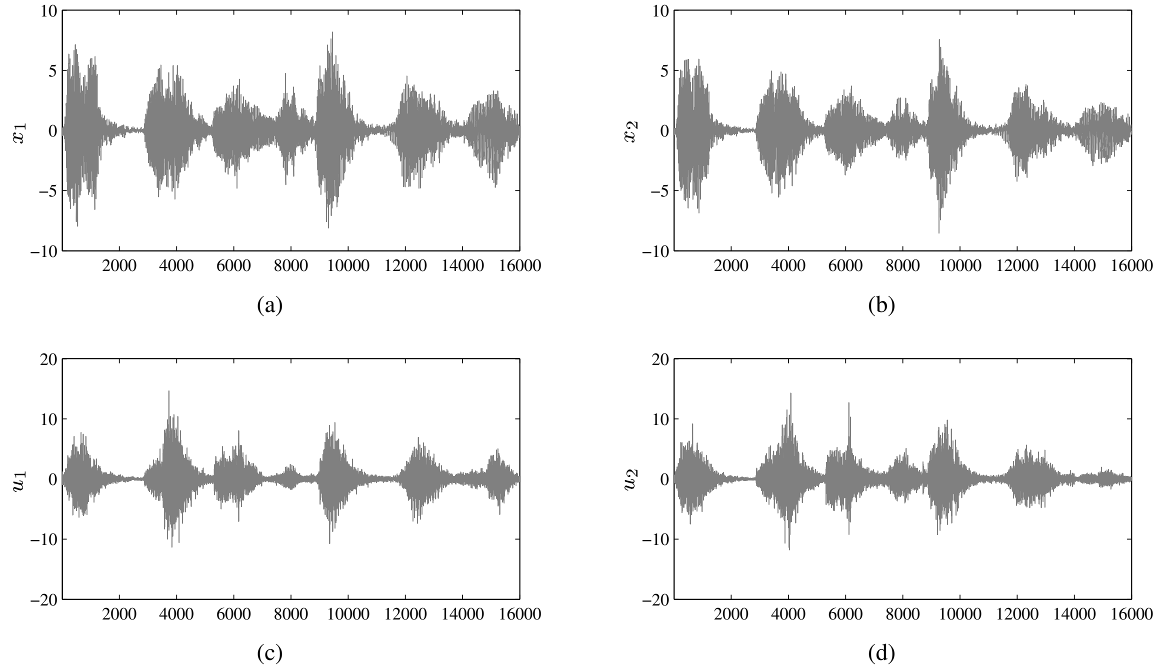


Figure 5.20 (a)–(b) Observable signals. (c)–(d) Estimated output signals.

5.7 Summary

Modelling of audio signals using the SαS and the NIG pdf.s are discussed, and it is demonstrated that the class of SαS distributions are attractive for modelling a broad class of audio signals.

A novel stochastic gradient adaptive filter that is suitable for non-Gaussian heavy tail signals is proposed. The adaptive filter is based on an empirical density transformation and it is demonstrated that the method has superior performance compared to state-of-the-art robust adaptive filtering methods. The applicability to audio signal processing is demonstrated in an acoustic echo canceling setup.

The independent component analysis methods based on joint diagonalization are examined for noisy mixtures of speech signals. It is demonstrated that the joint diagonalization methods can indentify the bases of speech signals, and the unified method in general has the best performance.

The independent component analysis method based on the spectral measure for αS signals are applied to an under-determined mixture of audio signals. Simulation examples demonstrate that the number of audio signals and the bases of the audio signals can be identified.

The maximum likelihood method for blind signal separation of convolutive mixtures is applied to a synthetic mixture of speech signals. The impulse response of the complete system is a diagonal system and thus signal separation has been obtained. However, due to the temporal dependency structure of the source signals there is a convolutive ambiguity. A frequency interpretation of the complete transfer function is provided and it appears that the convolutive ambiguity corresponds to a frequency weigthing of the estimated output signals. From a listening experiment it appears that the speech signals have not been completely separated, however, each of the speech signals appears to be considerably stronger in one channel than in the other.

The maximum likelihood method for blind signal separation of convolutive mixtures is applied to signals from an experimental teleconference system. It is demonstrated that the speech signals from simultaneous active speakers can be extracted from the microphone signals. The spatial responses of the separation system are determined, and the spatial responses are given an interpretation in terms of the location of the speakers.

CHAPTER 6

CONCLUSION

The basis of statistical signal processing from probability theory, mathematical statistics and information theory is reviewed, and it is demonstrated that second order statistics in general are insufficient for blind estimation methods. To overcome the limitations of second order statistics the theoretical foundation for moments, cumulants and non-linear estimation functions are established, and thereby the fundamentals for operating with heavy tail distributed signals and blind signal separation is established.

Several heavy tail distributions are introduced in a signal processing context, and special emphasis is on the class of αS distributions and the class of NIG distributions. The temporal dependence structure of infinite variance stochastic processes cannot directly be revealed by second order moment functions and therefore the fractional lower order moment function, the normalized moments and cumulants, and the covariation function is introduced. Methods for estimation of parameters in the SαS and NIG distributions suitable in signal processing are reviewed. The multivariate αS distributions are characterized by a spectral measure and a method for estimation of the spectral measure is introduced.

A probabilistic generative model is established for instantaneous mixtures of temporal independent sources, for instantaneous mixtures of sources with temporal dependence, and for convolutive mixtures of temporal independent sources. The maximum likelihood method for blind signal separation is derived for all three generative models.

The likelihood functions for SαS and NIG source distributions are examined, and it is shown that the likelihood function has maximum when the estimation matrix equals the inverse of the mixture matrix even for relatively large deviations between the model distribution and the true distribution of the sources. It is shown that the stochastic gradient method is useful in the maximum likelihood method for blind separation of heavy tail distributed source signals.

Several methods for independent component analysis based on joint diagonalization of estimated matrices are considered. A new method based on spatial covariation matrices that are suitable for SαS signals is proposed, and the theoretical basis for using the method based on spatial covariance matrices in the case of αS signals is established. A unifying method that combines any methods that can be formulated as a joint diagonalization problem is proposed. It is demonstrated that the unifying method in general has superior performance compared to other joint diagonalization methods.

A novel method for independent component analysis based on the discrete spectral measure for αS distributions is proposed. The method can identify the number of independent component and the non-orthogonal bases of the components. The method does not require a priori knowledge of the number of independent components, and the bases of the components can be identified even in the case of under-determined mixtures.

Modelling of audio signals using the SαS and the NIG distributions are discussed, and it is demonstrated that the class of SαS distributionss are attractive for modelling a broad class of audio signals.

A novel stochastic gradient adaptive filter that is suitable for non-Gaussian heavy tail signals is proposed. The adaptive filter is based on an empirical density transformation and it is demonstrated that the method has superior performance compared to state-of-the-art robust adaptive filtering methods. The applicability to audio signal processing is demonstrated in an acoustic echo canceling setup.

The independent component analysis methods based on joint diagonalization are examined for noisy mixtures of speech signals. It is demonstrated that the joint diagonalization methods can identify the bases of speech signals, and the unified method in general has the best performance.

The independent component analysis method based on the spectral measure for αS signals is applied to an under-determined mixture of audio signals. Simulation examples demonstrate that the number of audio signals and the bases of the audio signals can be identified.

The maximum likelihood method for blind signal separation of convolutive mixtures is applied to a synthetic mixture of speech signals. Due to the temporal dependency structure of the speech signals there is a convolutive ambiguity which corresponds to a frequency weigthing of the estimated signals. From a listening experiment it appears that the speech signals to some extent have been separated.

The maximum likelihood method for blind signal separation of convolutive mixtures is applied to signals from an experimental teleconference system. It is demonstrated that the speech signals from simultaneous active speakers can be extracted from the microphone signals. The spatial responses of the separation system is determined, and the spatial responses are given an interpretation in terms of the location of the speakers.

Research Contributions

The following list summarizes the research contributions established in the thesis.

- The basis for blind signal separation of heavy tail signals is established.
- An empirical density transformation that transforms a signal into an arbitrary distribution is proposed.
- The blind signal separation likelihood function is examined for SαS and NIG source distributions. It is shown that the stochastic gradient method is applicable in the maximum likelihood method for blind separation of heavy tail signals.
- An independent component analysis method based on spatial covariation matrices is proposed, and the theoretical basis for using the Molgedey-Schuster method in the case of αS signals is established.
- A method that unifies all independent component analysis methods based on joint diagonalization is proposed.
- An independent component analysis method based on the spectral measure is proposed. The method can identify the number of independent components and the bases of the components even for under-determined mixtures.
- It is demonstrated that a broad class of audio signals have heavy tail distributions, and that SαS distributions are suitable for modelling audio signals.
- A robust adaptive filtering method, based on an empirical density transformation, is proposed.
- It is demonstrated that the speech signals from simultaneously active speakers can be extracted from the microphone signals from a teleconference system. The estimated separation system is given a spatial interpretation that links the field of blind signal separation and beamforming.

Appendix A

Papers

This appendix includes the papers that have been published in relation to the work presented in the thesis.

[54] Preben Kidmose, Adaptive Filtering for Non-Gaussian Processes, *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pp. 424-427, 2000.

[55] Preben Kidmose, Alpha-Stable Distributions in Signal Processing of Audio Signals, *Proceedings of 41st Conference on Simulation and Modelling, Scandinavian Simulation Society*, pp. 87-94, 2000.

[35] Thomas Fabricius, Preben Kidmose and Lars Kai Hansen, Dynamic Components of Linear Stable Mixtures from Fractional Low Order Moments, *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pp. 3957-3960, 2001.

[56] Preben Kidmose, Independent Components Analysis Using the Spectral Measure for Alpha-Stable Distributions, *Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, pp. 1-5, 2001.

A.1 Adaptive Filtering for Non-Gaussian Processes

This section contains the paper *Adaptive Filtering for Non-Gaussian Processes*, [54]. The paper is published in the proceedings of *The 2000 IEEE International Conference on Acoustics, Speech and Signal Processing* pp. 424-427.

The paper proposes a new stochastic gradient robust filtering method based on a non-linear amplitude transformation. The method requires no a priori knowledge of the characteristics of the input signals and it is insensitive to the distribution and the stationarity of the signals. A simulation study, applying both synthetic and real-world signals, shows that the proposed method has overall better robustness performance, in terms of modelling error, compared with state-of-the-art robust filtering methods.

ADAPTIVE FILTERING FOR NON-GAUSSIAN PROCESSES

Preben Kidmose

Department of Mathematical Modelling, Section for Digital Signal Processing,
Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark

ABSTRACT

A new stochastic gradient robust filtering method, based on a non-linear amplitude transformation, is proposed. The method requires no a priori knowledge of the characteristics of the input signals and it is insensitive to the signals distribution and to the stationarity of the signals.

A simulation study, applying both synthetic and real-world signals, shows that the proposed method has overall better robustness performance, in terms of modeling error, compared with state-of-the-art robust filtering methods. A remarkable property of the proposed method is that it can handle double-talk in the acoustical echo-cancellation problem.

1. INTRODUCTION

Many signals encountered in practice are decidedly non-Gaussian and the signals are only stationary up to an order less than 2. Recently [6] has investigated audio signals, and demonstrated the usefulness of the α-Stable distribution for modelling of noisy audio signals.

In this work we consider robust adaptive algorithms. The objective is to design adaptive algorithms that are insensitive to the probability distribution and stationarity of the input signals.

1.1. Stochastic Gradient Algorithms

Consider an adaptive FIR filter with input u_n. The estimation error, at time n, of the filter is

$$e_n = d_n - \mathbf{w}^T \mathbf{u}_n \tag{1}$$

where $\mathbf{u}_n$ is the input signal vector, $\mathbf{w}$ is the filter coefficient vector, and d_n is the desired signal. Assume that d_n and u_n are jointly stable processes of order α. The objective is to minimize the dispersion of the error, $\min \left|\left|d_n - \mathbf{w}^T\mathbf{u}_n\right|\right|_\alpha$. This objective turns out to be intractable, but fortunately proportional to the pth order moment for any $0 < p < \alpha$, [7]. So, an equivalent objective is $J = \mathrm{E}\left\{\left|d_n - \mathbf{w}^T\mathbf{u}_n\right|^p\right\}$. There is no closed-form solution for the set of coefficients that minimize this objective; however the objective is convex for $1 \le p \le \alpha$, and so we may use a stochastic gradient method to solve for the coefficients. The class of stochastic gradient algorithms under consideration has the update

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu h(e_n \mathbf{u}_n) \tag{2}$$

where $h(\cdot)$ is an estimate of the gradient $\hat{\mathrm{E}}\left\{\partial J/\partial \mathbf{w}\right\}$, based on the estimation error and the input.

1.2. Adaptive Filters Based on FLOM

A number of different robust stochastic gradient algorithms, based on a fractional lower order moments (FLOM), has been proposed [1][2][3][5]. One of these is the Symmetric Least Mean P-norm (SLMP), that uses the symmetric norm

$$h(e\mathbf{u}) = (e\mathbf{u})^{\langle p/2 \rangle}$$

as estimate of the gradient, [3]. Another subclass of normalized stochastic gradient algorithms, that has the form of Eq. 2, is proposed in [5]. The proposed filter update uses

$$\left[h_q(\mathbf{u}_n)\right]_i = \frac{|u_{i,n}|^{q-1}\mathrm{sign}(u_{i,n})}{\sum_{m=1}^{L}|u_{i,n}|^q} \quad \text{for } 1 \le q < \infty \tag{3}$$

where $\left[h_q(\mathbf{u})\right]_i$ denotes the ith element of the vector valued function $h_q(\mathbf{u})$. This update, for $\mu = 1$ and for a valid q, corresponds to minimizing $\|\mathbf{w}_{n+1} - \mathbf{w}_n\|_p$ subject to $d_n - \mathbf{w}_{n+1}^T\mathbf{u}_n = 0$. The relation between the norm, q, of the input signal vector, $\mathbf{u}$, and the q-norm, of the coefficient vector update, $\|\mathbf{w_{n+1}} - \mathbf{w_n}\|_p$, is $1/p + 1/q = 1$, [5]. Thus, the update in Eq. 3 provides the minimum p-norm of the coefficient vector update. For $q = 2$ the algorithm is the classical NLMS algorithm.

Recently [1] proposed a generalization to the NLMP with the update

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{e_n^{\langle a \rangle}}{\|\mathbf{u}_n\|_{qa}^{qa} + \lambda}\mathbf{u}_n^{\langle (q-1)a \rangle} \tag{4}$$

where, in the presence of an α-stable process, $0 < a \le \alpha - 1$ and $1 \le q$. For any real number v and $p \ge 0$, the convention $v^{\langle p \rangle} = |v|^p \mathrm{sign}(v)$ is used. The update is motivated by observing that it corresponds to a gradient descent adaptation approach to the objective function $J = \mathrm{E}\left\{|e_n|^{a+1}\right\}$. The update in Eq. 4 reduces to the NLMP update if a and q are chosen as $p-1$ and p/a. The update in equation 3 is a special case of Eq. 4 for $a = 1$ and $\lambda = 0$

1.3. Median Orthogonality

If the median of the product, $M(u_1 u_2)$, of two random variables u_1 and u_2, is zero, then u_1 and u_2 are said to be median orthogonal, $u_1 \perp_{\mathrm{MO}} u_2$. For random variables with symmetric probability densities, independence is necessary and sufficient for MO, however MO is necessary but not sufficient for independence. The MO filter criterion, proposed in [2], is that the error should be MO to all elements of the input vector $e \perp_{\mathrm{MO}} \mathbf{u}$. This criterion extends the conventional orthogonality criterion without restricting the distribution of e and $\mathbf{u}$.

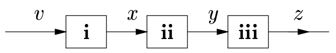
2. A NEW ROBUST FILTERING METHOD

In this section we propose a conceptual different approach to robust adaptive algorithms, that are based on a non-linear amplitude transformation. The idea of the amplitude transformation is to ensure the existence of the second order moment.

2.1. Density Transformation

The purpose of the non-linear amplitude transformation, is to transform the probability density of a signal. The proposed transformation, from an arbitrary unknown density to a desired density, is provided by a three-step procedure

- **i** Use the empirical density transformation, to obtain an uniform density.
- **ii** Transform with the inverse function to the indefinite integral of the desired density.
- **iii** Scale the transformed variable.



The scaling is necessary since all scale information is lost in the empirical density transformation. The objective of the scaling is to obtain $\|v\|_p = \|z\|_p$, for a sufficient low p.

2.2. Empirical Transformation to Uniform Density

Let $\tilde{v}$ be a random variable, and let $v_1, v_2, \ldots, v_N$ be N observations of the variable. An empirical transformation from the unknown arbitrary probability density random variable, $\tilde{v}$, to the variable $\tilde{x}$ with uniform density, can be done by sorting the observations. Sort the N observations, by the index m, such that $v_{\cdot,1} \leq v_{\cdot,2} \leq \cdots \leq v_{n,m} \leq \cdots \leq v_{\cdot,N}$. The variable $\tilde{x}$ is forced to follow the uniform density

$$p(x) = \begin{cases} \frac{1}{2}r & \text{if} \quad |x-a| \leq r \\ 0 & \text{if} \quad |x-a| > r \end{cases}$$

by assigning x_n the value

$$x_n = (2m - Nr)/N + a$$

where m is the index that corresponds to the nth observation of v.

The transformation is non-parametric and can be interpreted as the inverse cumulated histogram. The transformation, independent of the observations, result in an variable with uniform density. For variables with median equal to zero the transformation is sign conserving, and for symmetric densities the transformation is an odd function. Dependencies between variables are conserved.

2.3. Deterministic Density Transformation

The probability density of a random variable $\tilde{x}$ is denoted $p(x)$. The relation between the probability density for the random variable $\tilde{x}$ and $\tilde{y}$, is determined by the fundamental law of probability, $|p(y)dy| = |p(x)dx|$. Consider a transformation of the random variable $\tilde{x}$, following the probability density $p(x)$, to a new variable $\tilde{y}$, following the probability density $p(y)$. Suppose that $\tilde{x}$ follow a uniform distribution, and let $p(y)$ be a positive function with integral equal 1. The desired transformation, that transform a uniform distributed variable $\tilde{x}$ into a variable $\tilde{y}$, that follows the density $p(y)$, is $y = g(x)$, where $g(x)$ is the inverse function to the indefinite integral of $p(y)$. The transformation from a uniform distributed variable $\tilde{x}$ into a variable $\tilde{y}$ that follow the probability density $p(y)$, is depicted in Fig. 1. Thus, transformation between an arbitrary continuous probability density to or from an uniform probability density, is given by the probability distribution function or the inverse probability distribution function respectively.

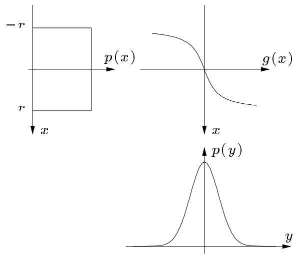


Figure 1: Density transformation from an uniform probability density, $p(x)$, to a probability density, $p(y)$, by the function $g(x)$.

2.4. Stochastic Gradient Adaptive Filter Based on non-linear Transformation

The idea of the non-linear amplitude transformation, is to make a density transformation, that ensures the existence of the second order moment. We propose to use a transformation, $g(\cdot)$, that transform an arbitrary density into a Normal distribution. If the estimation error, e_n, and the input signal vector, $\mathbf{u}_n$, is transformed into a Normal distributions by the function $g_e(\cdot)$ and $g_{\mathbf{u}}(\cdot)$, then the objective $\mathrm{E}\left(\left|g\left(d_n - \mathbf{w}^T\mathbf{u}_n\right)\right|^2\right)$ makes sense.

The structure of the adaptive algorithm is depicted in Fig. 2. The proposed amplitude transformation is sign conserving, and that ensures that $g_{\mathbf{u}}(\mathbf{u})\, g_e(e)$ is an ascending direction for the objective function. A straightforward update equation, on the form of Eq. 2, is

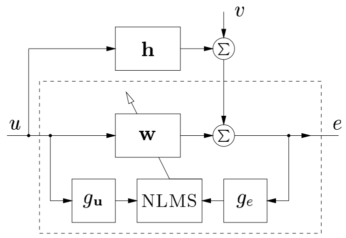


Figure 2: Adaptive filter based on a non-linear transformation of the estimation error.

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu g_{\mathbf{u}}(\mathbf{u}_n)\, g_e(e_n) \tag{5}$$

The median orthogonality (MO) criterion proposed in [2] apply to this update. The algorithm will have a stability point where $e \perp_{MO} \mathbf{u}$ because the integral of the even density function and the odd $g(\cdot)$ is zero.

3. SIMULATIONS

The setup used for evaluation is depicted in Fig. 2, where $\mathbf{h}$ is an 10th order low pass filter. The proposed algorithm is compared to 3 other algorithms, that represents state of the art in robust adaptive algorithms:

$$\begin{aligned}
&\text{NSLMP} && \mathbf{w}_{n+1} &&= \mathbf{w}_n - \mu \frac{1}{\|\mathbf{u}\|_p^p + a}(e\mathbf{u})^{\langle p/2 \rangle} \\
&\text{NLMP} && \mathbf{w}_{n+1} &&= \mathbf{w}_n - \mu \frac{1}{\|\mathbf{u}\|_p^p + a} e^{\langle p-1 \rangle}\mathbf{u} \\
&\text{Aydíns method} && \mathbf{w}_{n+1} &&= \mathbf{w}_n - \mu \frac{\mathbf{u}^{\langle (q-1)a \rangle}}{\|\mathbf{u}\|_{qa}^{qa} + \lambda} e^{\langle a \rangle}
\end{aligned}$$

The performance of the algorithms are evaluated as the modeling error $10 \cdot \log_{10} \mathrm{E}\left\{(\mathbf{w}-\mathbf{h})^T(\mathbf{w}-\mathbf{h})/(\mathbf{h}^T\mathbf{h})\right\}$.

3.1. Simulation with synthetic signals

The synthetic signals are modeled as sequences of independent symmetric α-stable distributed random variables. The characteristic function for the random variable is given by $\phi(t) = \exp(-\gamma|t|^\alpha)$. The simulation of the signals are based on the method proposed in [4],[7].

The adaptation constants used in the simulation study, are listed in Table 1. Obviously these constants have considerably influence on the characteristics of the algorithms. The constants are chosen such that all the algorithms are robust and stable in the wide range of signals under consideration in this simulation study. Note that, in the advantage of the algorithms used for comparison, the norms used in these algorithms are adjusted in accordance with the applied signals.

Algorithm	Adaptation constants
Aydíns method	$\mu = .01$, $p = q = 1.1$, $a = \alpha_v - 1.1$
NSLMP	$\mu = .01$, $p = .9\alpha_v$, $a = 10^{-6}$
NLMP	$\mu = .01$, $p = .9\alpha_v$, $a = 10^{-6}$
Proposed method	$\mu = .05$, $N = 120$

Table 1: Adaptation constants used in the system identification simulation study with α-stable signals

Consider the case where the input signal, u, and the interference signal, v, are identical white Cauchy distributed signals. The convergence of the modeling error is depicted in Fig. 3. All the methods converge to about the same misadjustment level, but the proposed method has the lowest misadjustment. Aydíns method has slower convergence than the other algorithms.

For the case where the source signal, u, is a white Gaussian signal, and the interfering signal, v, is white following a Cauchy distribution, the convergence of the modeling error is depicted in Fig. 4. Two important changes are observed: the misadjustment level is about 20 dB higher and the convergence speed is slower. The γ-ratio is not a suitable measure between variables of different distributions, so even though the γ-ratio between the signals is still equal to one, the empirical SNR is very poor.

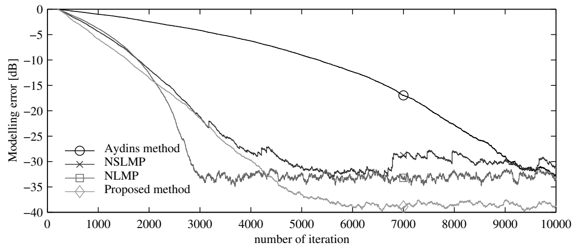


Figure 3: Modeling error convergence for $\alpha_u = \alpha_v = 1$ and $\gamma_u = \gamma_v = 1$. Mean over 10 Monte Carlo runs.

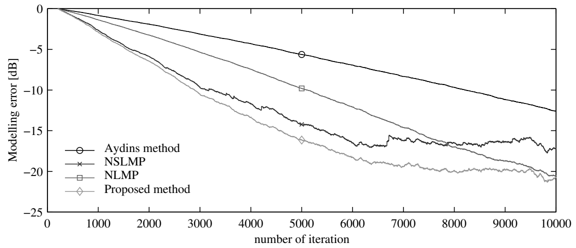


Figure 4: Modeling error convergence for $\alpha_u = 2$, $\alpha_v = 1$ and $\gamma_u = \gamma_v = 1$. Mean over 10 Monte Carlo runs.

In Fig. 5 the modeling error versus signal to interference dispersion ratio is depicted. The dispersion of the source signal is varied in the interval $\gamma_u = [0.1, 100]$, and for the interfering signal the dispersion is constant, $\gamma_v = 1$. For low dispersion ratios the performance of the algorithms are much the same; for increasing dispersion ratio the algorithms has increasing performance, and the proposed method takes increasing advantages compared to the other algorithms.

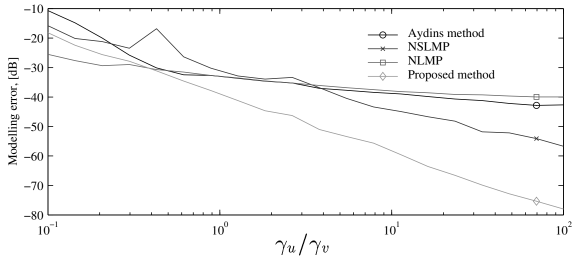


Figure 5: Modeling error versus signal to interference dispersion ratio, $\frac{\gamma_v}{\gamma_u}$. Characteristic exponent $\alpha_u = \alpha_v = 1$.

In Fig. 6 the modeling error versus the characteristic exponent is depicted. The methods used for comparison are almost independent of α, since in this simulation study the norm is adjusted in accordance with the applied signal. Even though the proposed method does not require any knowledge of the input signals distribution it outperforms all the algorithms used for comparison. This is a remarkably robustness characteristic.

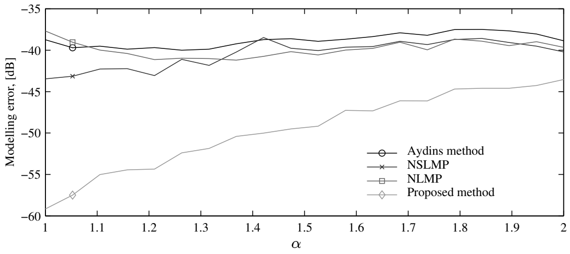


Figure 6: Modeling error versus the characteristic exponent, α. The dispersion is $\gamma_u = 1$ and $\gamma_v = 0.1$. Mean over 10 Monte Carlo runs.

3.2. Simulation with speech signals

Consider the adaptive algorithms used for acoustical echo canceling. The setup is as depicted in Fig. 2 where v is the local speaker and u is the remote speaker. The acoustical path, modeled by the filter $\mathbf{h}$, is the same 10th order filter as in the previous simulation study. The adaptation constants used in this example are listed in Table 2. For the NSLMP, NLMP and Aydíns method the p-norm and μ are set to relative small values to obtain stability. The signals are speech

Algorithm	Adaptation constants
Aydíns method	$\mu = .01$, $p = q = 1.1$, $a = .1$
NSLMP	$\mu = .01$, $p = 1.44$, $a = 10^{-6}$
NLMP	$\mu = .01$, $p = 1.44$, $a = 10^{-6}$
Proposed method	$\mu = .1$, $N = 1000$

Table 2: Adaptation constants used in the acoustical echo cancellation for speech signals

signals sampled at 8kHz. The input signals and the modeling error result of the simulation is shown in Fig. 7. The

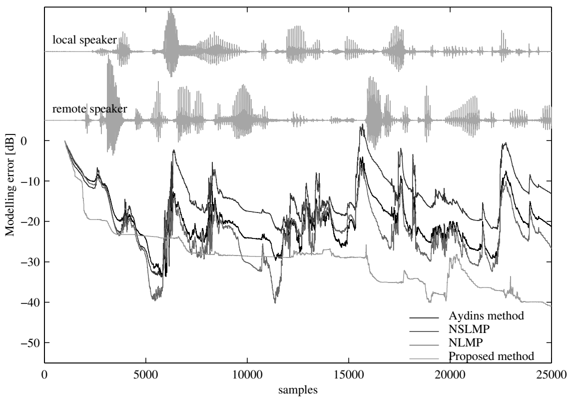


Figure 7: Modeling error, in acoustical echo-canceler in cross-talk environment, for the 4 methods.

simulation example contains several difficulties: the signals are non-stationary and contains temporal correlation, the amplitude density is non-Gaussian, and the dispersion ratio between the signals are strongly varying. It appears, from Fig. 7, that the proposed method has slower convergence but apart from that, considerably better performance compared with the other methods. In the intervals where the local speaker is active, which corresponds to a poor SNR, the algorithms, except for the proposed method, loss the tracking of the filter. For these methods a control strategy is necessary. Contrary to the methods used for comparison, the proposed method has the ability to handle double-talk situations and environments with high background noise levels.

4. CONCLUSION

A new robust adaptive filtering method, based on a nonlinear amplitude transformation, is proposed. The method is compared with state-of-the-art robust filtering methods and a simulation study with synthetic and real-world signals is carried out.

The proposed method requires no a priori knowledge of the input signals, it is insensitive to the signals density and to the stationarity of the signals.

The proposed method has an overall better robustness performance, compared with state-of-the-art methods. The method is computational more expensive than the algorithms used for comparison, but these methods require either the use of a low p-norm or the estimation of a valid p, in order to ensure stability.

A remarkable byproduct is that the proposed amplitude transformation method can be used in a variety of situations for signal conditioning.

5. REFERENCES

[1] Gül Aydín, Orban Aríkan, and A. Enis Çetin. Robust Adaptive Filtering Algorithms for α-Stable Processes. *IEEE transactions on Circuits and Systems - II*, 46(2):198–202, February 1999.

[2] John S. Bodenschatz and Chrysostomos L. Nikias. Recursive Local Orthogonal Filtering. *IEEE Transactions on Signal Processing*, 45(9):2293–2300, 1997.

[3] John S. Bodenschatz and Chrysostomos L. Nikias. Symmetric Alpha-Stable Filter Theory. *IEEE Transactions on Signal Processing*, 45(9):2301–2306, 1997.

[4] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[5] Scott C. Douglas. A Family of Normalized LMS Algorithms. *IEEE Signal Processing Letters*, 1(3):49–51, March 1994.

[6] Panayiotis G. Georgiou, Panagiotis Tsakalides, and Chris Kyriakakis. Alpha-Stable Modeling of Noise and Robust Time-Delay Estimation in the Presence of Impulsive Noise. *IEEE Transactions on Multimedia*, 1(3):291–301, September 1999.

[7] Chrysostomos L. Nikias and Min Shao. *Signal Processing with Alpha-Stable Distribution and Applications.* Wiley, 1 edition, 1995.

A.2 Alpha-Stable Distributions in Signal Processing of Audio Signals

This section contains the paper *Alpha-Stable Distributions in Signal Processing of Audio Signals*, [55]. The paper is published in the proceedings of *The 41st Simulation Conference, Scandinavian Simulation Society* pp. 87-94.

The main contributions of this paper is

1. The proposal of two versions of a sliding window, block based parameter estimator for estimating the parameters in a symmetrical stable distribution. The proposed estimator is suitable for parameter estimation in audio signals.

2. Discussion of the suitability of the stable distribution, for modelling audio signals. For a broad class of audio signals, the distribution and the stationarity property are examined. Empirically it is shown that the class of stable distributions provides a better model for audio signals, than the Gaussian distribution model.

For demonstrating the applicability of stable distributions in audio processing, a classical problem from statistical signal processing, stochastic gradient adaptive filtering, is considered.

Alpha-Stable Distributions in Signal Processing of Audio Signals

Preben Kidmose,
Department of Mathematical Modelling, Section for Digital Signal Processing,
Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark

Abstract

First, we propose two versions of a sliding window, block based parameter estimator for estimating the parameters in a symmetrical stable distribution. The proposed estimator is suitable for parameter estimation in audio signals.

Second, the suitability of the stable distribution, for modelling audio signals, is discussed. For a broad class of audio signals, the distribution and the stationarity property are examinated. It is empirically shown that the class of stable distributions provides a better model for audio signals, than the Gaussian distribution model.

Third, for demonstrating the applicability of stable distributions in audio processing, a classical problem from statistical signal processing, stochastic gradient adaptive filtering, is considered.

1 Introduction

The probability density of many physical phenomena have tails that are heavier than the tails of the Gaussian density. If a physical process has heavier tails than the Gaussian density, and if the process has the probabilistic stability property, the class of stable distributions may provide a useful model.

Stable laws have found applications in diverse fields, including physics, astronomy, biology and electrical engineering. But despite the fact that the stable distribution is a direct generalization of the popular Gaussian distribution, and shares a lot of its useful properties, the stable laws have been given little attention from researchers in signal processing.

A central part of statistical signal processing, is the linear theory of stochastic processes. For second order processes, the theory is established, and numerous algorithms are developed. Applying these algorithms on lower order processes, results in considerably performance degradation or the algorithms may not even be stable. Thus, there is a need for developing algorithms based on linear theory for stable processes, and these algorithms could improve performance and robustness.

2 Modelling Audio Signals

The class of stable distributions is an appealing class for modelling phenomena of impulsive nature; and it is to some extent analytical tractable, because of two important properties: it is a closed class of distributions, and it satisfies the *generalized central limit theorem*.

Audio signals in general are not stationary, the temporal correlation is time varying, and it turns out that the probability density function is more heavy tailed than a Gaussian density. In this work we assume that probability density is symmetric, which is a weak restriction for audio signals. In particular we will consider the Symmetric Alpha Stable, SαS, distribution [5]. To show the general applicability of the SαS for modelling audio signals, we examine six audio signals, with very different characteristics.

2.1 The SαS Distribution

A univariate distribution function is SαS if the characteristic function has the form

$$\phi(t) = \exp(-\gamma|t|^{\alpha})$$

where the real parameters α and γ satisfies $0 < \alpha \leq 2$ and $\gamma > 0$. The parameter α is called the characteristic exponent. The smaller α-value the more probability mass in the tails of the density function. For $\alpha = 2$ it is the Gaussian distribution, and for $\alpha = 1$ it is the Cauchy distribution. The scale parameter, γ, is denoted the dispersion.

For stable distributions, moments only exist for order less than the charateristic exponent. If x is a SαS random variable, then the fractional lower order moment is

$$\mathrm{E}\left(|x|^p\right) = C(p,\alpha)\gamma^{\frac{p}{\alpha}} \quad 0 < p < \alpha$$

where $C(p,\alpha) = \left(2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{-p}{\alpha}\right)\right) / \left(\alpha\sqrt{\pi}\Gamma\left(\frac{-p}{2}\right)\right)$ and $\Gamma(\cdot)$ is the usual Γ-function [5, 4].

2.2 Parameter Estimation for SαS Distribution

Several methods, for estimation of the parameters in stable distributions, have been proposed in the literature, see [3, 6] and references herein. In this work, the estimation is performed in the $\ln|\mathrm{S}\alpha S|$-process, as proposed in [3]. Estimation in the $\ln|\mathrm{S}\alpha S|$-process has reasonable estimation characteristics, low computational complexity, and the estimator is on a closed form expression. For a SαS random variable x, the estimation in the $\ln|\mathrm{S}\alpha\mathrm{S}|$-process is given as

$$\sigma_z^2 = \frac{\pi^2}{6}\left(\frac{1}{\alpha_x^2} + \frac{1}{2}\right) \quad (1)$$

$$\mu_z = C_e\left(\frac{1}{\alpha_x} - 1\right) + \frac{1}{\alpha_x}\ln\gamma_x \quad (2)$$

where μ_z and σ_z^2 is the first and second moment of $z = \ln|x|$, and $C_e = 0.57721566\ldots$ is the Euler constant. The characteristics of the α-estimator is depicted in Fig.1. The estimator in Eq.1 requires many

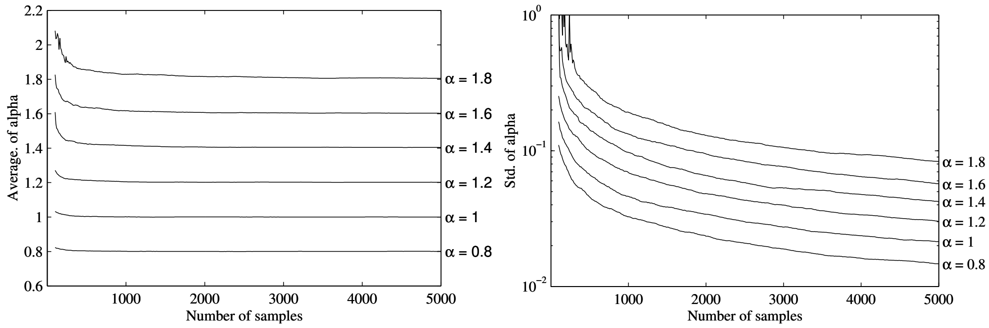


Figure 1 Characteristics of the α-estimator in Eq.1. Average (left) and standard deviation (right) of the α-estimate versus number of samples in estimation.

samples to give low variance estimates, and the variance is dependent of the characteristic exponent.

2.3 Sliding Window, Block Based Parameter Estimation

In this section we propose two versions of a sliding window, block based parameter estimator, suitable for audio signals. The basic idea of the estimator, is based on two observations: Audio signals often has strong short term correlations, due to mechanical resonans. These short term correlations have big influence on the short term distribution, this is particularly the case for mechanical systems with low damping combined with heavy tailed exitation signals. And the stationarity characteristics of the audio signals, necessitate the use of a windowed parameter estimation.

Thus, there is a need for a windowed estimator, that is robust to the influence from short terms correlations. Basically the proposed estimator has two steps, that makes the estimator suitable for handling these characteristics: A short term decorrelation, based on a linear prediction filter. And a sliding window, block based updating of the SαS-parameters of the decorrelated signal sequences. The short term decorrelation is performed over a block of M samples, and the updating of the SαS-parameters is performed over N blocks.

Let k denote the current block, the decorrelation is then performed over $x(n)|_{n=(k-1)M+1,\ldots,kM}$, and the total window length, MN samples, applies over $x(n)|_{n=(k-N)M+1,\ldots,kM}$.

Mechanical resonans is well modelled by a simple low order AR system, thus the resonans part of the signal can be removed by a linear predictor. The linear predictor coefficients for the kth block, $a_k(l)$, $l = 1, \ldots, L$, is determined by

$$\sum_{l=1}^{L} a_k(l) r_x(m-l) = r_x(m)$$

where $r_x(m)$ is the autocorrelation sequence of $x(n)|_{n=(k-1)M+1,\ldots,kM}$. The decorrelated signal, $y(n)$, in the kth block, is determined as the inverse filter

$$y(n)|_{n=(k-1)M+1,\ldots,kM} = x(n) - \sum_{l=1}^{L} a_k(l) x(n-l)$$

In order to apply the estimator in the $\ln|\mathrm{S}\alpha\mathrm{S}|$-process, we define the signal $z(n) = \ln|y(n)|$. A windowed estimator of $\mathrm{E}(z)$ over N blocks of M samples is

$$\hat{\mu}_z(k) = \frac{1}{NM} \sum_{m}^{kM} z(m)$$

where the sum is over the N blocks, i.e. $m = (k-N)M + 1, \ldots, kM$. The estimator of the expectation value is updated as

$$\hat{\mu}_z(k) = \hat{\mu}_z(k-1) - \frac{1}{N} \sum_{m}^{(k-N)} z(m) + \frac{1}{N} \sum_{m}^{k} z(m) \tag{3}$$

where the first sum is over $m = (k-N-1)M + 1, \ldots, (k-N)M$, and the second sum is over $m = (k-1)M + 1, \ldots, kM$. Similar a windowed estimator of $\mathrm{E}\left(z(m) - \mathrm{E}(z(m))\right)^2$, over N blocks, is

$$\begin{aligned}
\hat{\sigma}_z(k) &= \frac{1}{NM} \sum^{k} \left(z(m) - \hat{\mu}_z(k)\right)^2 \\
&= \hat{\sigma}_z(k-1) - \frac{1}{N}\left(\sum^{k-N} z^2(m) - \sum^{k} z^2(m)\right) + \hat{\mu}_z^2(k-1) - \hat{\mu}_z^2(k) \qquad (4)
\end{aligned}$$

Note that if the block sum and the quadratic block sum in Eq.3 and Eq.4 is saved for the last N blocks, then it is only necessary to calculate the kth block sum and the kth quadratic block sum for each block update. The windowed, block based estimators for the first and second moment of $z(n)$, combined with the $\mathrm{S}\alpha\mathrm{S}$ parameter estimators in Eq.1 and 2, yields[1]:

$$\begin{aligned}
\hat{\alpha}(k) &= \left(\frac{6}{\pi^2}\hat{\sigma}_z^2(k) - \frac{1}{2}\right)^{-1/2} \qquad (5)\\
\hat{\gamma}(k) &= \exp\left(\left(\hat{\mu}_z(k) - \left(\frac{1}{\hat{\alpha}(k)} - 1\right) C_e\right)\hat{\alpha}(k)\right) \qquad (6)
\end{aligned}$$

The dynamic properties of estimators are of crucial importance in the case of non-stationary signals. The α-estimator in Eq.5 is sensitive to abrupt changes in the distribution, which might be an undesirable properties for non-stationary signals. An estimator, that is more robust against abrupt changes in distribution, is

$$\hat{\alpha}(k) = \frac{1}{N}\sum_{n=0}^{N-1}\left(\frac{6}{\pi^2}\left(\frac{1}{M}\sum_{m=0}^{M-1}\left(z(k-m-nM) - \frac{1}{M}\sum_{l=0}^{M-1} z(k-l-nM)\right)^2\right) - \frac{1}{2}\right)^{-1/2} \tag{7}$$

which is the empirical mean over N blocks of length M of the α-estimator in Eq. 1. This estimator is biased. In Fig.2 the proposed estimators is applied on a block-wise stationary $\mathrm{S}\alpha\mathrm{S}$ signal, and it is apparent that the estimator in Eq.7 is more robust in the case of abrupt changes in distribution.

[1] If the update in Eq.3 and 4 is modified to be an iterative block based update, the update equations in Eq.5 and 6 is equivalent to the iterative update equations proposed in [3]. However, in this context, it is the sliding window property that is the important feature.

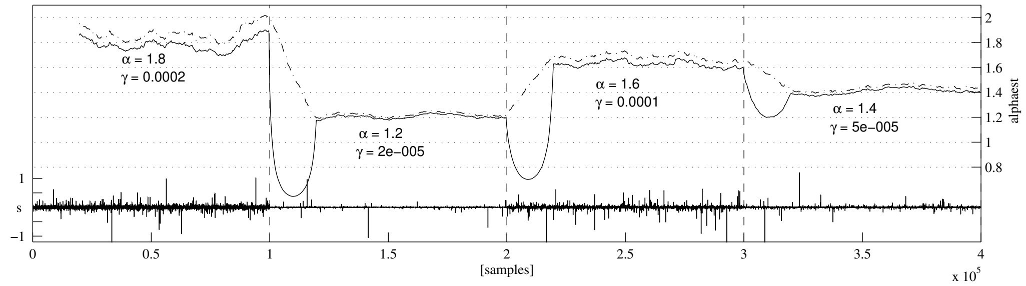


Figure 2 Dynamic properties of the proposed α-estimators. The solid line is the estimator in Eq.5, the dashed-dot line is the estimator in Eq.7. The signal, s, is stationary block-wise over 100000 samples; the estimator block size $M = 400$ and the sliding window is over $N = 50$ blocks.

2.4 SαS Modelling of Audio Signal

The signals used for demonstrating the applicability of SαS modelling, are depicted in the bottom part of the plots in Fig. 3, and some additional informations are listed in Tabel 1. In Fig.4 the empirical, the estimated Gausssian, and the estimated SαS density, for the signal sequences, are depicted. These density plots are obtained over the whole signal length. It is important to notice that, due to the non-stationarity

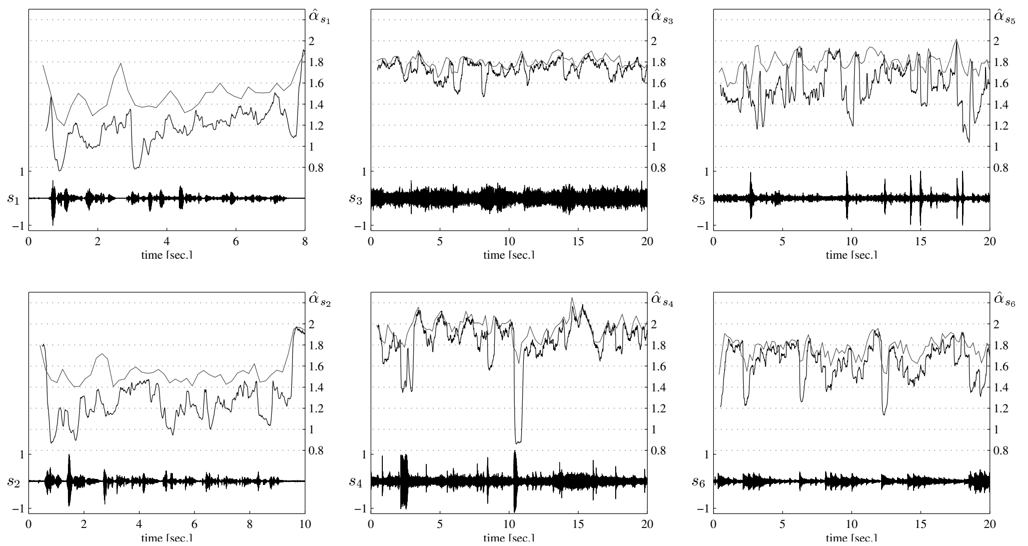


Figure 3 Estimates of the characteristic exponent, α. Solid lines: estimator Eq.5, the linear predictor window is 10 ms and $N = 50$ blocks. Dotted lines: estimator Eq.7, the linear predictor window is over 100 ms, and the mean is over $M = 5$ blocks. For both estimators the linear predictor filter is of order 12.

of the signals, the time window, in which the density is estimated, has deciding influence of the density estimate. Comparing the Gaussian model, which has only one degree of freedom in modelling of the shape of the probability density, with the SαS distribution, which has two degrees of freedom, it is reasonable to conclude that the SαS in general provides a better model for the probability density.

It is instructive to consider the distribution of the signals in shorter time windows. The proposed sliding window estimators in Eq.5 and Eq.7 are applied to the six signals, and the estimate of the characteristic exponent is depicted on the same time axis as the signals in Fig 3. The solid line is the characteristic exponent estimated with the parameter estimator in Eq.5. The linear predictor window is 10 ms, and the estimator window is over 50 blocks. Due to the different sampling frequencies 20 kHz, 32 kHz, and 44.1 kHz, this corresponds to a window length of 200, 320 and 441 samples respectively. The dotted lines is the

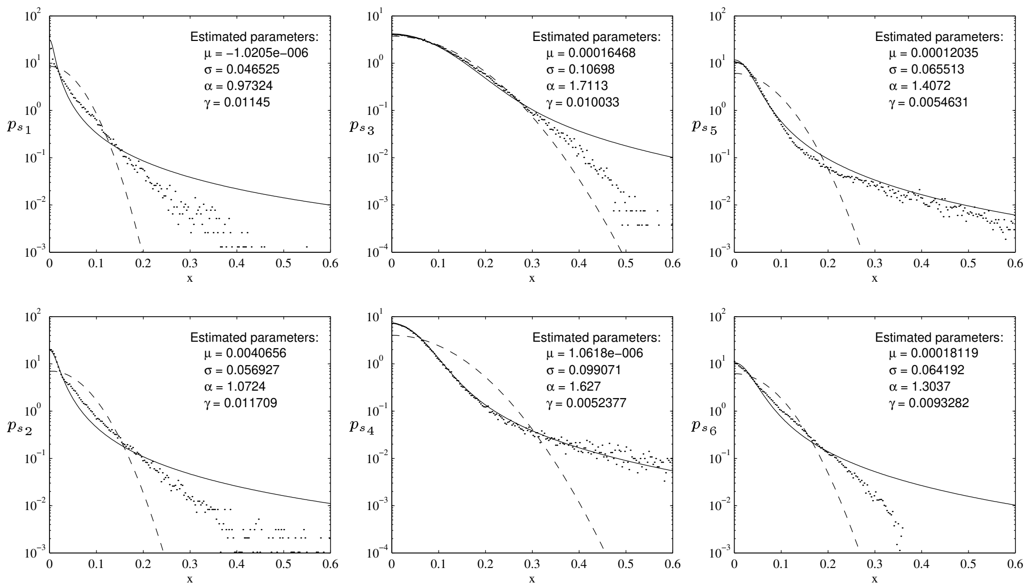


Figure 4 Probability density functions for the signals. The dots indicates the empirical density function. The dased line is the Gaussian density corresponding to the estimated mean, μ, and standard deviation, σ^2. The solid line is the SαS density corresponding to the estimated characteristic exponent, α, and dispersion, γ. The estimated parameter values, for each signal, is tabulated in the upper right corner.

characteristic exponent estimated with the parameter estimator in Eq.7. The linear predictor window is over 100 ms, and the mean is over 5 blocks. For both estimators the linear predictor filter is of order 12.

The short term estimates of the characteristic exponent are in general larger than the long term estimates. It is interesting to notice that the speech signals has relative low α-values and in certain intervals approaches the Cauchy distribution. The background noise signals has α-values that in general are considerable larger, but still far below the Gaussian distribution. In Fig.5 the density estimates are depicted for the signal $s_2(n)$ and $s_4(n)$ for two different time windows, and again it is reasonable to conclude that the SαS provide a better fit to the empirical histogram.

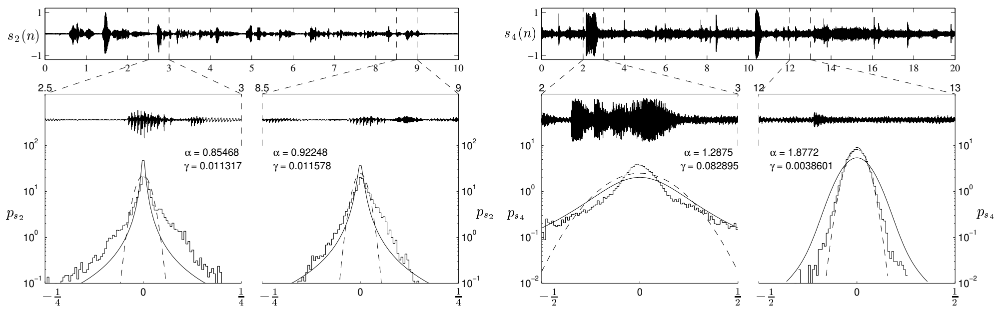


Figure 5 Examples of short term density estimates for the signals $s_2(n)$ and $s_4(n)$ for different time intervals. The stair plot indicates the empirical density function. The dased line is the Gaussian density corresponding to the estimated mean and standard deviation. The solid line is the SαS density corresponding to the estimated characteristic exponent, α, and dispersion, γ. The block size is 320 and 200 samples respectively which corresponds to 10ms. The estimator in Eq.5 is applied over 50 and 100 blocks respectively, and no linear prediction filter has been used.

It is well-known from the theory of stable distributions, that moments only exists for moments of order less than α. The preceeding examinations, that indicates that the SαS is suitable for modelling audio signals, and that the characteristic exponent varies between a Cauchy and a Gaussian distribution, exposes the idea to use the estimates of the characteristic exponent in variable fractional lover order moment adaptive

Signal	Describtion
$s_1(n)$	Speech signal, male, low background noise. Sampling freq. 32 kHz., 8 sec.
$s_2(n)$	Speech signal, male, low background noise. Sampling freq. 32 kHz., 10 sec.
$s_3(n)$	Cocktail party background noise. Sampling freq. 44.1 kHz., 20 sec.
$s_4(n)$	Background noise recorded in a kitchen. Sampling freq. 20 kHz., 20 sec.
$s_5(n)$	Background noise recorded in an office. Sampling freq. 44.1 kHz., 20 sec.
$s_6(n)$	Music, classical guitar. Sampling freq. 44.1 kHz., 20 sec.

Table 1 Additional information for the audio signals.

algorithms. This idea is the issue of the following section.

3 Adaptive Filtering

An illustrative application for adaptive filtering, is the acoustical echo canceller. The objective is to cancel out the loudspeaker signal from the microphone signal, see Fig.6. An adaptive filter is applied to estimate the acoustical channel from the loudspeaker to the microphone. The echo cancelled signal is obtained by subtracting the remote signal, filtered by the estimated acoustical channel, from the microphone signal. From an algorithmic point of view the local speaker is a noise signal, and the applied adaptiv algortihm must exhibit adequate robustnes against this noise signal.

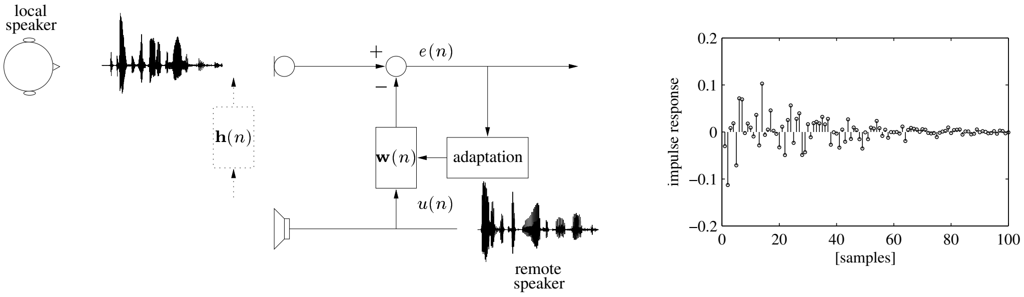


Figure 6 Left: Acoustical echo canceller setup. Right: Impulse response for the acoustical channel, h, from loudspeaker to microphone.

The standard algorithm for adaptive filters, is the Normalized Least Mean Square (NLMS) algorithm, with the update

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{\mathbf{u}(n)e(n)}{a + ||\mathbf{u}(n)||_2^2}$$

The NLMS algorithm has severe convergence problems for signals with more probability mass in the tails, than the Gaussian distribution. Recently filter theory for SαS signals has been developed [1], and the Least Mean P-norm (LMP) algorithm is proposed. The LMP algorithm is significant more robust to signal with heavy tails. In the following simulation study a normalized LMP update is applied:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{\mathbf{u}(n)e(n)^{p-1}}{a + ||\mathbf{u}(n)||_p^p}$$

For SαS signals the p-norm must be less than the characteristic exponent, $p < \alpha$. The development of robust adaptive filters is subject to active research, see [2] and references herein.

Consider the adaptiv echo canceller setup, with the signal scenario as depicted in the upper part of Fig.7. The local speaker is the speech signal s_1, the speaker is deactivated in the time interval 1-2 sec, and added some cocktail party background noise. The remote speaker is the signal s_2, the speaker is deactivated in the time interval 4-6 sec. and in that time interval the loadspeaker signal is a Gaussian noise signal (comfort noise, $\sigma^2 = .005$). The α-estimator in Eq.5 is applied to the error signal, $e(n)$, the block length is 640, the estimator runs over 10 blocks, and the linear predictor filter is of order 12. Three different adaptive

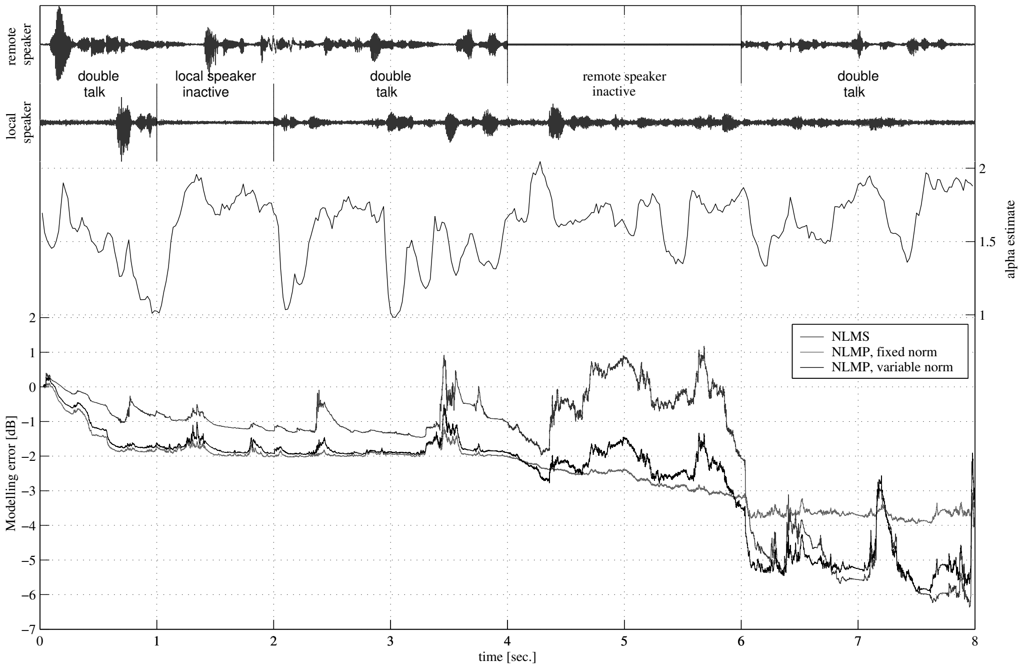


Figure 7 Simulation result for acoustical echo canceller scenario.

filters is applied, the standard NLMS, the NLMP with a fixed norm ($p = 1.1$), and the NLMP with a norm that is adjusted in accordance with the α-estimate ($p = \hat{\alpha} - .2$). The stepsize parameter $\mu = .001$ for NLMS, and $\mu = .002$ for NLMP. The performance of the algortihms is evaluated as the modelling error $10 \cdot \log_{10} \mathrm{E}\left\{(\mathbf{w} - \mathbf{h})^T(\mathbf{w} - \mathbf{h})/(\mathbf{h}^T\mathbf{h})\right\}$, and is depicted in the bottom part of Fig.7.

The $p = 1.1$ for the fixed norm NLMP was empirical found to be the best norm for the applied signals, and the NLMP algorithm with fixed norm in general performed very good. The modelling error for NLMP algorithm with variable norm, is between the modelling error of the two others algorithms. The variable norm algorithm follows the best of the two other algorithms, and it is thus concluded that the variable norm algortihm has overall better performance. However, the result is far from convincing, and the conclusion might not hold in general, because of the dependence of the α-estimator.

The variable norm NLMP algorithm is computational much more expensive than using a fixed norm algorithm, because of the running α-estimator. The small gain in performance do probably not justify the additional computational expenses. Despite this fact, the simulation study shows, that the choice of norm has deciding influence of the performance of the algorithms.

4 Conclusion

The proposed sliding window, block based parameter estimators has been applied to a broad class of audio signals. Comparing the histogram of the audio signals with the estimated parameters of a SαS distribution, it is concluded that the class of SαS distributions is suitable for modelling audio signals.

The simulation study shows that lower norms algorithms exhibit better robustness characteristic for audio signals, and that the choice of norm has deciding influence of the performance of the algorithm.

Stable distributions provides a framework for synthesis of robust algortihms for a broad class of signals. The linear theory of stable distributions and processes, and the development of robust algorithms for impulsive signals is an open research area.

References

[1] John S. Bodenschatz and Chrysostomos L. Nikias. Symmetric Alpha-Stable Filter Theory. *IEEE Transactions on Signal Processing*, 45(9):2301–2306, 1997.

[2] Preben Kidmose. Adaptive Filtering for Non-Gaussian Processes. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pages 424–427, 2000.

[3] Xinyu Ma and Chrysostomos L. Nikias. Parameter Estimation and Blind Channel Identification in Impulsive Signal Environments. *IEEE Transactions on Signal Processing*, 43(12):2884–2897, December 1995.

[4] Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman&Hall, 1994.

[5] Min Shao and Chrysostomos L. Nikias. Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications. *Proceeding of the IEEE*, 81(7):986–1010, July 1993.

[6] George A. Tsihrintzis and Chrysostomos L. Nikias. Fast Estimation of the Parameters of Alpha-Stable Impulsive Interference. *IEEE Transactions on Signal Processing*, 44(6):1492–1503, June 1996.

A.3 Dynamic Components of Linear Stable Mixtures from Fractional Low Order Moments

This section contains the paper *Dynamic Components of Linear Stable Mixtures from Fractional Low Order Moments*, [35]. The paper is published in the proceedings of *The 2001 IEEE International Conference on Acoustics, Speech and Signal Processing* pp. 3957-3960.

The second moment based independent component analysis scheme of Molgedey and Schuster is generalized to fractional low order moments, relevant for linear mixtures of heavy tail stable processes. The Molgedey and Schuster algorithm is based on spatial covariance matrices and it allows explicitly construction of the independent components. This turns out to be possible also for spatial fractional low order moment matrices.

DYNAMIC COMPONENTS OF LINEAR STABLE MIXTURES FROM FRACTIONAL LOW ORDER MOMENTS

Thomas Fabricius, Preben Kidmose and Lars Kai Hansen

Technical University of Denmark
Department for Mathematical Modelling
Richard Petersens Plads bldg. 321
2800 Lyngby
Denmark

ABSTRACT

The second moment based independent component analysis scheme of Molgedey and Schuster is generalized to fractional low order moments, relevant for linear mixtures of heavy tail stable processes. The Molgedey and Schuster algorithm stands out by allowing explicitly construction of the independent components. Surpricingly, this turns out to be possible also for decorrelation based on fractional low order moments.

Keywords: Blind Source Separation (BSS), Dynamic Components, Independent Component Analysis (ICA), Stable process, Fractional lower order moments, Cocktail party problem

1. INTRODUCTION

Reconstruction of independent components from linear mixtures is an important signal processing research area with numerous practical applications [1, 2]. Typically, independent component analysis (ICA) proceeds from non-linear transformations [3, 4] or from temporal correlations [5, 6]. The second moment based Molgedey-Schuster approach assumes that the independent sources have different autocorrelation functions [5, 7, 8]. The main virtue of the approach is that an explicit solution (reconstruction of sources and mixing matrix) is possible. In [8] we applied the Molgedey-Schuster algorithm to image mixtures and we proposed a minor modification of the algorithm that relieves a problem of the original approach, namely that it occasionally produces complex mixing coefficients and source signals. Attias and Schreiner proposed a very rich ICA framework based on higher order statistics and decorrelation [9, 10], allowing for completely general and learnable source distributions, however at the price of significant additional computation.

In this contribution we analyze dynamic decorrelation for heavy tail stable random signal, and we provide a generalization of the Molgedey-Schuster algorithm based on fractional low order moments rather than second moments.

This work is funded by the Danish Research Councils through the THOR Center for Neuroinformatics.

2. STABLE DISTRIBUTIONS AND SOME PROPERTIES

Random vectors following a stable distribution are stable with respect to addition, i.e., if $\mathbf{X}$ and $\mathbf{Y}$, are stable then so is the sum $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$. A random vector follows a strictly symmetric stable distribution *iff* its characteristic function has the form [11]

$$\phi(\mathbf{t}) = \begin{cases} \exp\left\{\mathbf{t}^T\mathbf{A}\mathbf{t}\right\} & \alpha = 2 \\ \exp\left\{\int_S \left|\mathbf{t}^T\mathbf{s}\right|^\alpha \mu(dS)\right\} & 0 < \alpha < 2 \end{cases} \tag{1}$$

where $\mathbf{t} \in R^n$ and $\mu(dS)$ is a finite Borel measure called the spectral measure, S is the unit sphere and α is the characteristic exponent. A random variable X that is distributed by a symmetric stable distribution with characteristic exponent α is denoted $X \sim S\alpha S$. In the class of stable distributions we find the normal distribution $\alpha = 2$ and the so-called Cauchy distribution $\alpha = 1$. In general the low order stable distributions have heavy tails, and are supposed to model, e.g., speech signals. Because of the heavy tails, higher order moments do not exists, hence, ICA methods based on second moments are not well-defined for these signals. For multivariate stable variables the spectral measure carries information about the dependencies among the individual components and this may be used for independent component analysis [12]. Here we show that independent dynamic components can be reconstructed explicitly *a la* Molgedey Schuster, even if the signals follow stable distributions.

Our derivation is based on the generalized covariation between random variables X and Y jointly $S\alpha S$ as defined in [11]

$$[X,Y]_\alpha = \int_S xy^{\langle\alpha-1\rangle}\mu(dS) \tag{2}$$

with the definition $z^{\langle\alpha-1\rangle} = |z|^{\alpha-1}\,\mathrm{sign}(z)$. Note that $[X,Y]_\alpha = 0$ is necessary but not sufficient for X and Y to be independent. The covariation has a convenient pseudo-linearity property. Let X_1, X_2, Y_1 and Y_2 be jointly $S\alpha S$ and Y_1, Y_2 independent then

$$\begin{aligned}[a_1X_1+a_2X_2,b_1Y_1+b_2Y_2]_\alpha = \\ a_1[X_1,Y_1]_\alpha b_1^{\alpha-1} + \quad a_2[X_2,Y_1]_\alpha b_1^{\alpha-1} \quad + \\ a_1[X_1,Y_2]_\alpha b_2^{\alpha-1} + \quad a_2[X_2,Y_2]_\alpha b_2^{\alpha-1}\end{aligned} \tag{3}$$

From the covariation we can derive the *covariation coefficient*

$$\lambda_{X,Y} = \frac{[X,Y]_\alpha}{[Y,Y]_\alpha} \tag{4}$$

which has the property [11]

$$\lambda_{X,Y} = \frac{E\left(XY^{\langle p-1\rangle}\right)}{E\left(YY^{\langle p-1\rangle}\right)} \tag{5}$$

for $1 \le p < \alpha$. By equation (5) we can estimate the covariation coefficient from data.

3. DECORRELATION BASED ON FRACTIONAL LOW ORDER MOMENTS

We are now equipped to generalize the Molgedey-Schuster approach, and start out by defining the matrix of observed signals $\mathbf{X}$ with elements

$$X_{m,t} = \textstyle\sum_{j=1}^{N_s} A_{m,j}S_{j,t} \quad 0 \le t < T \tag{6}$$

where $A_{m,j}$ are the mixing coefficients forming the mixing matrix $\mathbf{A}$, $S_{j,t}$ is an instance of the j^{th} $S\alpha S$ source signal written as an element in the matrix $\mathbf{S}$ and N_s is the number of sources. We now assume that the different source signals are independent,

$$[S_{j,t}, S_{i,t+\tau}]_\alpha = 0 \quad j \neq i \quad \tau \in R. \tag{7}$$

The covariation coefficient between $X_{m,t}$ and $X_{n,t+\tau}$ can be found using equations (4) and (6)

$$\lambda_{X_{m,t},X_{n,t+\tau}} = \frac{\left[\sum_{j=1}^{N_s} A_{m,j}S_{j,t}, \sum_{i=1}^{N_s} A_{n,i}S_{i,t+\tau}\right]_\alpha}{\left[\sum_{i=1}^{N_s} A_{n,i}S_{i,t+\tau}, \sum_{i=1}^{N_s} A_{n,i}S_{i,t+\tau}\right]_\alpha} \tag{8}$$

Finally, we use that the sources are independent, hence, equation (3) is valid, providing

$$\lambda_{X_{m,t},X_{n,t+\tau}} = \frac{\sum_{j=1}^{N_s}\sum_{i=1}^{N_s} A_{m,j}\,[S_{j,t},S_{i,t+\tau}]_\alpha\, A_{n,i}^{\langle\alpha-1\rangle}}{\sum_{j=1}^{N_s}\sum_{i=1}^{N_s} A_{n,j}\,[S_{j,t+\tau},S_{i,t+\tau}]_\alpha\, A_{n,i}^{\langle\alpha-1\rangle}} \tag{9}$$

We define matrices $\mathbf{\Lambda}^{(\tau)}$ with elements $\Lambda_{m,n}^{(\tau)} = \lambda_{X_{m,t},X_{n,t+\tau}}$, $\mathbf{B}^{(\tau)}$ with the nominator of equation (8) as elements

$$B_{m,n}^{(\tau)} = \sum_{j=1}^{N_s}\sum_{i=1}^{N_s} A_{m,j}\,[S_{j,t},S_{i,t+\tau}]_\alpha\, A_{n,i}^{\langle\alpha-1\rangle} \tag{10}$$

and a diagonal matrix $\mathbf{D}$ with the denominator from equation (8) as elements

$$D_{n,n} = \sum_{j=1}^{N_s}\sum_{i=1}^{N_s} A_{n,j}\,[S_{j,t+\tau},S_{i,t+\tau}]_\alpha\, A_{n,i}^{\langle\alpha-1\rangle} \tag{11}$$

Assuming stationarity, $\mathbf{D}$ is independent of τ, so that $\mathbf{\Lambda}^{(\tau)} = \mathbf{B}^{(\tau)}\mathbf{D}^{-1}$. From equation (10) it is seen that $\mathbf{B}^{(\tau)}$ decomposes into

$$\mathbf{B}^{(\tau)} = \mathbf{A}\mathbf{C}^{(\tau)}\mathbf{A}^{T\,\langle\alpha-1\rangle} \tag{12}$$

where $\mathbf{A}^{T\,\langle\alpha-1\rangle}$ is the signed norm applied to each element of $\mathbf{A}^T$. $\mathbf{C}^{(\tau)}$ is the matrix with elements $C_{j,i}^{(\tau)} = [S_{j,t}, S_{i,t+\tau}]_\alpha$, and we notice that $\mathbf{C}^{(\tau)}$ is diagonal since the sources are independent. We next form the quotient matrix $\mathbf{Q}^{(\tau)}$,

$$\begin{aligned}\mathbf{Q}^{(\tau)} &= \mathbf{\Lambda}^{(\tau)}\mathbf{\Lambda}^{(0)^{-1}} = \mathbf{B}^{(\tau)}\mathbf{B}^{(0)^{-1}} \\ &= \mathbf{A}\mathbf{C}^{(\tau)}\mathbf{A}^{T\,\langle\alpha-1\rangle}\mathbf{A}^{T\,\langle\alpha-1\rangle^{-1}}\mathbf{C}^{(0)^{-1}}\mathbf{A}^{-1} \\ &= \mathbf{A}\mathbf{C}^{(\tau)}\mathbf{C}^{(0)^{-1}}\mathbf{A}^{-1}\end{aligned} \tag{13}$$

We can now construct the mixing matrix $\mathbf{A}$ by solving the eigenvalue problem,

$$\mathbf{Q}^{(\tau)}\mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Gamma} \tag{14}$$

if we identify $\mathbf{\Psi} = \mathbf{A}$ and $\mathbf{\Gamma} = \mathbf{C}^{(\tau)}\mathbf{C}^{(0)^{-1}}$.

In short, we have shown our main result, namely that it is possible to recover the mixing matrix from fractional low order moments. The source signals are subsequently estimated by $\mathbf{S} = \mathbf{A}^{-1}\mathbf{X}$.

3.1. Sample estimates of Fractional Low Order Moments

Equation (5) provides the estimate of the elements of $\mathbf{\Lambda}^{(\tau)}$

$$\widehat{\lambda}_{X_{m,t},X_{n,t+\tau}} = \frac{\frac{1}{T-\tau}\sum_{t=0}^{T-1-\tau} X_{m,t}X_{n,t+\tau}^{\langle p-1\rangle}}{\frac{1}{T}\sum_{t=0}^{T-1} X_{n,t}X_{n,t}^{\langle p-1\rangle}} \tag{15}$$

from a sample of T measurements. We can then form

$$\begin{aligned}\widehat{\mathbf{Q}}^{(\tau)} &= \widehat{\mathbf{\Lambda}}^{(\tau)}\,\widehat{\mathbf{\Lambda}}^{(0)\,-1} \\ &= \frac{T}{T-\tau}\mathbf{X}\mathbf{X}^{(\tau)\langle p-1\rangle\,T}\left(\mathbf{X}\mathbf{X}^{\langle p-1\rangle\,T}\right)^{-1} \quad (16)\end{aligned}$$

hence, solve the eigenvalue problem to obtain the estimated mixing matrix. In the above equation $\mathbf{X}^{(\tau)}$ is the zero padded matrix $\mathbf{X}$ shifted by a lag τ. The choice of τ is a question one have to deal with. The optimal choice will both depend on the accuracy on the estimated elements in $\mathbf{C}^{(\tau)}$ and $\mathbf{C}^{(0)}$ and on the actual true values in $\mathbf{C}^{(\tau)}\mathbf{C}^{(0)\,-1}$. τ should be selected so we have the largest difference in the eigenvalues of $\mathbf{Q}^{(\tau)}$ i.e. in $\mathbf{C}^{(\tau)}\mathbf{C}^{(0)\,-1}$ compared to their accuracy. In this paper we will not address this further, but suggest using $\tau = 1$.

4. EXPERIMENTAL RESULTS

We illustrate the performance of the generalization of the Molgedey-Schuster algorithm by analyzing mixing of synthetic data following a stable distribution, and by mixing of two speech signals.

4.1. Synthetic data

We perform experiments on mixing of two signals in two channels. In the first we mix two independent unit-variance Gaussian signals by the mixing matrix

$$\mathbf{A} = \begin{bmatrix} 0.8944 & 0.4472 \\ 0.3714 & 0.9285 \end{bmatrix}$$

containing two non-orthogonal normalized rows. The source signals have different temporal dependencies, one is white noise, while the other is low-pass filtered with a FIR-filter with two coefficients $h_0 = 1$ and $h_1 = 0.8$. We show in Figure 1 scatterplots of the "true signals" and the recovered signals from the estimated mixing matrix estimated from the conventional Molgedey-Schuster algorithm. Next we mix two similar stable signals drawn from distributions with $\alpha = 1.5$, hence with heavy tails, by the same mixing matrix. In Figure 2 we show the scatterplots of the true and estimated signals.

The two different scenarios lead to similar estimated mixing matrices

$$\mathbf{A} = \begin{bmatrix} 0.9034 & 0.4287 \\ 0.2812 & 0.9597 \end{bmatrix}; \quad \mathbf{A} = \begin{bmatrix} 0.9367 & 0.3503 \\ 0.4272 & 0.9041 \end{bmatrix}$$

demonstrating that it is possible reliably to reconstruct independent stable source signals by the generalized Molgedey-Schuster scheme, in cases where the second moment correlations would not converge for large samples.

4.2. Speech data

Finally, we consider the blind separation of a mixture of two speech signals. The signals are two source signals used in [13]. We use the same mixing matrix as in the above example. In Figure 3 we show the scatterplot of the true and estimated signals using the conventional Molgedey-Schuster algorithm and in Figure 4 using the new scheme using $p = 1.8$. The recovered signals are very similar, indicating that the conventional algorithm is robust to the underlying distribution of the source signals. We conjecture that this robustness is related to the fact that the conventional Molgedey-Schuster algorithm proceeds from the ratio of two second order moments, and that the ratio might have better statistical properties than the two individual moments, this is the topic for on-going research.

5. CONCLUSION

We have derived a new algorithm generalizing that of Molgedey and Schuster [5] based on fractional low order moments. We have provided a scheme for estimating the relevant moments from finite samples and demonstrated the viability of the new algorithm in a synthetic data set. However, we also found that the original algorithm is fairly robust to possibly diverging second order moments, and conjectured that the explanation could be related to the fact that the conventional algorithm is based on the ratio between two diverging terms, possibly improving the statistical properties.

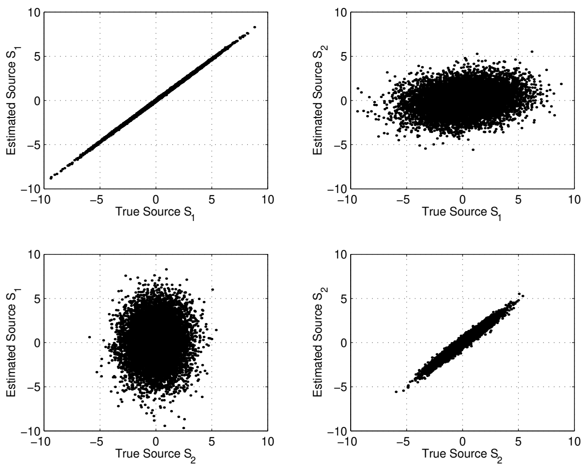


Fig. 1. Original Gaussian source, i.e. $\alpha = 2$ signals against recovered source signals when using the original Molgedey and Schuster algorithm ($p = 2$). In this setup $T = 10000$ and one of the sources was white, the other filtered by the FIR-filter in the text. The algorithm was run at $\tau = 1$

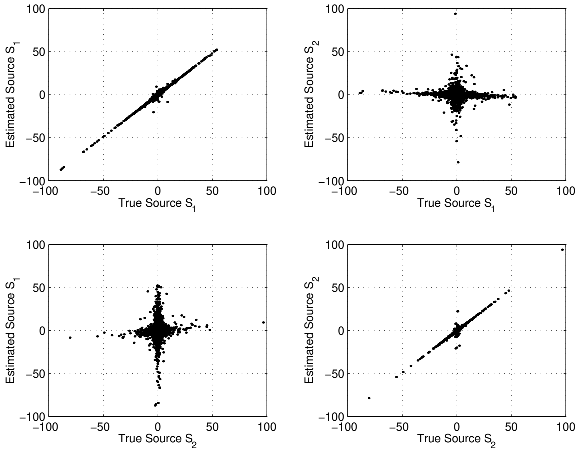


Fig. 2. Same setup as in Fig.1 but now the source signals are generated with α=1.5 and the algorithm is run with p=1.49.

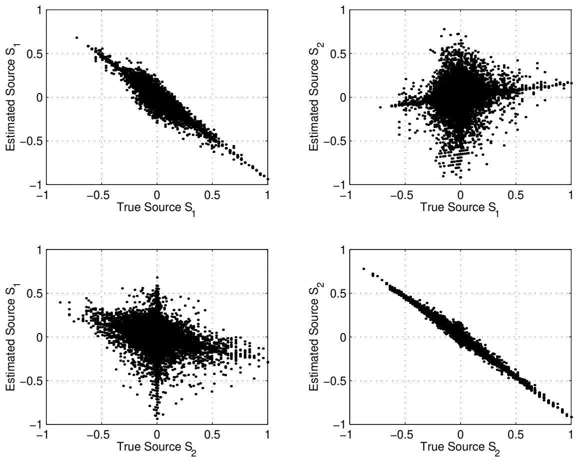


Fig. 3. Original speech source signals against recovered source signals when using the original Molgedey and Schuster algorithm. The speech signals had a duration of $T = 23000$. The algorithm was run at $\tau = 5$.

6. REFERENCES

[1] Te-Won Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, New York, 1998.

[2] M. Girolami (Ed.), *Advances in Independent Component Analysis.*, Springer-Verlag, New York, 2000.

[3] P. Comon, "Independent component analysis: A new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[4] Anthony J. Bell and Terrence J. Sejnowski, "An Information-Maximisation Approach to Blind Separation and Blind De-

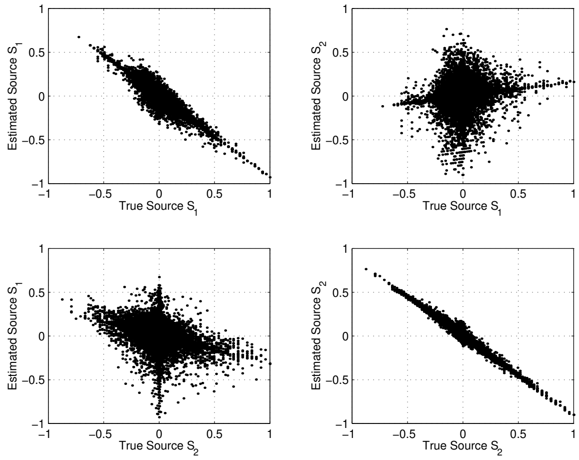


Fig. 4. The setup is the same as that in Fig.3 but with $p = 1.8$

convolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[5] L. Molgedey and H. Schuster, "Separation of independent signals using time-delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.

[6] A.Belouchrani, K.Abed-Meraim, J.F.Cardoso, and E.Moulines, "Blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[7] L.K.Hansen and J.Larsen, "Source separation in short image sequences using delayed correlation," *Proceedings of the IEEE Nordic Signal Processing Symposium, Vigsø, Denmark 1998. Eds. P. Dalsgaard and S.H. Jensen*, pp. 253–256, 1998.

[8] L.K.Hansen, J.Larsen, and T.Kolenda, "On independent component analysis for multimedia signals," *L. Guan, S.Y. Kung and J. Larsen (eds.) Multimedia Image and Video Processing, CRC Press*, 2000.

[9] H. Attias and C.E. Schreiner, "Blind source separation and deconvolution by dynamic component analysis," *Neural Networks for Signal Processing VII: Proceedings of the 1997 IEEE Workshop, 456-465 (1997*, pp. 456–465, 1997.

[10] H. Attias and C.E. Schreiner, "Blind Source Separation and Deconvolution: The Dynamic Component Analysis Algorithm," *Neural Computation*, vol. 10, pp. 1373–1424, 1998.

[11] Chrysostomos L. Nikias and Min Shao, *Signal Processing with Alpha-Stable Distribution and Applications*, Wiley, 1 edition, 1995.

[12] P.Kidmose, "Independent components of stable linear mixtures," 2001, Submitted to ICASSP2001.

[13] T.W.Lee, M.S.Lewicki, M.Girolami, and T.J.Sejnowski, "Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations," *IEEE Signal Processing Letters*, vol. 6, pp. 87–90, April 1999.

A.4 ICA Using the Spectral Measure for Alpha-Stable Distributions

This section contains the paper *Independent Components Analysis based on the Discrete Spectral Measure*, [56]. The paper is published in the proceedings of the *IEEE - EURASIP, Workshop on Nonlinear Signal and Image Processing, 2001*.

The paper proposes a method for identifying the independent components of a stable random vector. The characteristic function of a stable vector is determined by a spectral measure, and the method is based on the observation that the spectral measure is discrete if the vector can be expressed as a linear combination of independent components. The spectral measure is determined based on an estimate of the characteristic function, and by evaluating the spectral measure the non-orthogonal bases of the mixture matrix are identified. A remarkable property of the method proposed in this paper is that both the number of independent component and the bases of a general rectangular matrix can be identified. Sampling properties of the characteristic function is discussed, and simulation examples demonstrates the method.

INDEPENDENT COMPONENT ANALYSIS USING THE SPECTRAL MEASURE FOR ALPHA-STABLE DISTRIBUTIONS

*Preben Kidmose**

Informatics and Mathematical Modelling, Richard Petersens Plads, Building 321,
Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark.

ABSTRACT

A method for identifying the independent components of an alpha-stable random vector is proposed. The method is based on an estimate of the spectral measure for the characteristic function. Simulations with both synthetic and speech signals demonstrate that the proposed method can identify independent components in the so-called over-complete case of more sources than mixtures.

Keywords: Alpha-stable distributions, spectral measure, independent component analysis, blind identification, overcomplete representation.

1. INTRODUCTION

Independent Component Analysis (ICA) addresses the problem of reconstruction of N sources from the observation of M instantaneous linear combinations thereof.

Stable distributions is suitable for modelling random variables with tails of the probability density function that are *heavier* than the tails of the Gaussian density function. Stable distributions have found applications in signal processing [1], and also in processing of audio signals [2, 3].

This work consider the ICA on stable vectors, and the paper is organized as follows. Stable distributions are introduced in Sec.2 and it is shown that a fundamental property of stable random vectors with independent components is that the so-called spectral measure is discrete (established result). In Sec.3 the estimation of the spectral measure is discussed, and in Sec.4 the spectral measure is used for identifying the number of independent components and the non-orthogonal bases of these components. The results of a simulation study, of both synthetic and speech signals, are presented in Sec.5.

2. STABLE DISTRIBUTIONS

Closed-form expressions do not in general exist for Alpha-Stable (αS) density and distribution functions, therefore the αS distributions are usually described by characteristic functions (ch.f.).

2.1. Characteristic Function for αS Variable

The ch.f. of a random variable x in $\mathbb{R}$ is defined as $\Phi(t) = \mathrm{E}\ \exp(\mathrm{itx})$, where E denotes the expectation. The random variable is αS with characteristic exponent α, if the ch.f. can be written on the form

$$\Phi(t) = \exp\left(-I(t) + i\mu t\right) \tag{1}$$

where

$$I(t) = \begin{cases} \gamma|t|^{\alpha}(1 - i\beta\tan(\dfrac{\alpha\pi}{2})\mathrm{sign}(t)) & , \quad \alpha \neq 1 \\ \gamma|t|(1 + i\beta\dfrac{\pi}{2}\mathrm{sign}(t)\ln(t)) & , \quad \alpha = 1 \end{cases}$$

and $0 < \alpha \leq 2$, $\gamma > 0$, $-1 \leq \beta \leq 1$, and $(\alpha, \beta, \gamma, \mu) \in \mathbb{R}$. The parameter β determines the symmetry of the distribution, γ is the scale parameter, and μ is the location parameter. For $\beta = 0$ and $\mu = 0$ the distribution is symmetric and is denoted $\mathrm{S}\alpha\mathrm{S}(\alpha, \gamma)$.

2.2. Characteristic Function for αS Vector

The ch.f. for a real random vector $\mathbf{x}$ is defined as $\Phi(\mathbf{t}) = \mathrm{E}\exp(i\mathbf{t}^T\mathbf{x})$. A random vector is an αS vector if a finite measure Γ exists on the unit sphere S_M of $\mathbb{R}^M$ such that the ch.f. can be written on the form

$$\Phi(\mathbf{t}) = \exp\left(-I(\mathbf{t}) + i\mathbf{t}^T\boldsymbol{\mu}\right) \tag{2}$$

where

$$I(\mathbf{t}) = \int_{S_M} \psi_\alpha(\mathbf{t}^T\mathbf{s})\, d\Gamma(\mathbf{s})$$

and

$$\psi_\alpha(u) = \begin{cases} |u|^{\alpha}(1 - i\tan(\frac{\alpha\pi}{2})\mathrm{sign}(u)) & , \quad \alpha \neq 1 \\ |u|(1 + i\frac{\pi}{2}\mathrm{sign}(u)\ln(u)) & , \quad \alpha = 1 \end{cases}$$

The measure Γ is called the *spectral measure* of the αS random vector, and $\boldsymbol{\mu}$ is called the shift vector. The spectral measure is a real function, and the spectral representation $(\Gamma, \boldsymbol{\mu})$ is unique. See [4, 1].

*This work is supported by the Danish Research Councils through the THOR Center for Neuroinformatics.

2.3. αS Vector with Discrete Spectral Measure

Consider the real vector $\mathbf{v} = [v_1 \dots v_N]^T$, where $v_n, n = 1, \dots, N$ are independent $\alpha\mathrm{S}(\alpha, \beta_n, \gamma_n, \mu_n)$ random variables. Let $\mathbf{x} = [x_1 \dots x_M]^T$ be a random vector in $\mathbb{R}^M$, and $\mathbf{A}$ be a $\mathbb{R}^{M\times N}$ matrix. The vector $\mathbf{x} = \mathbf{A}\mathbf{v}$ is then an αS random vector with ch.f.

$$\Phi(\mathbf{t}) = \mathrm{E}\exp\left(i\mathbf{t}^T\mathbf{A}\mathbf{v}\right) = \prod_{n=1}^{N} \mathrm{E}\exp\left(i\mathbf{t}^T\mathbf{a}_{\cdot n}v_n\right)$$

where $\mathbf{a}_{\cdot n}$ is the nth column of $\mathbf{A}$. This is essentially the product of all the ch.f.s of v_n, and can be written as

$$\Phi(\mathbf{t}) = \exp\left(-I_N(\mathbf{t}^T\mathbf{a}_{\cdot n}) + i\sum_{n=1}^{N}\mu_n\mathbf{t}^T\mathbf{a}_{\cdot n}\right) \quad (3)$$

where

$$I_N(\mathbf{t}^T\mathbf{a}_{\cdot n}) = \sum_{n=1}^{N}\gamma_n\psi_\alpha\left(\mathbf{t}^T\mathbf{a}_{\cdot n}\right) \quad (4)$$

Rewriting (3) to the form of (2) yields

$$\begin{aligned}\Gamma(\mathbf{s}) = \; & \sum_{n=1}^{N}\frac{1+\beta_n}{2}\gamma_n(\mathbf{a}_{\cdot n}^T\mathbf{a}_{\cdot n})^{\alpha/2}\delta(\mathbf{s}-\mathbf{s}_n) + \\ & \sum_{n=1}^{N}\frac{1-\beta_n}{2}\gamma_n(\mathbf{a}_{\cdot n}^T\mathbf{a}_{\cdot n})^{\alpha/2}\delta(\mathbf{s}+\mathbf{s}_n)\end{aligned} \quad (5)$$

where $\mathbf{s}_n = \mathbf{a}_{\cdot n}/(\mathbf{a}_{\cdot n}^T\mathbf{a}_{\cdot n})^{1/2}$ and

$$\boldsymbol{\mu} = \begin{cases}\displaystyle\sum_{n=1}^{N}\mathbf{a}_{\cdot n}\mu_n & , \quad \alpha \neq 1 \\ \displaystyle\sum_{n=1}^{N}\mathbf{a}_{\cdot n}\left(\mu_n - \frac{1}{\pi}\sigma_n\beta_n\ln(\mathbf{a}_{\cdot n}^T\mathbf{a}_{\cdot n})\right) & , \quad \alpha = 1\end{cases}$$

See [4, Example 2.3.6]. Hence the spectral measure $\Gamma(\mathbf{s})$ of the αS random vector $\mathbf{x}$ is discrete and concentrated on N symmetric pairs of points $(\mathbf{s}_n, -\mathbf{s}_n)$, $n = 1, \dots, N$ of S_N. This result holds in general, thus the spectral measure $\Gamma(\mathbf{s})$ of an αS random vector $\mathbf{x}$ is discrete on the unit sphere S_N if, and only if, the αS random vector can be expressed as a linear transformation of independent αS random variables.

3. ESTIMATION OF THE SPECTRAL MEASURE

We describe a method for estimation of the spectral measure, examine the statistical characteristics of the estimator, and discus the sampling radius of the ch.f..

3.1. Discrete Approximation of the Spectral Measure

From [5, Theorem 1] we have that for a d-dimensional αS variable with spectral measure $\Gamma(\mathbf{s})$ and density function $p(\mathbf{x})$, there is a discrete spectral measure $\Gamma_a(\mathbf{s})$ with corresponding density function $p_a(\mathbf{x})$ satisfying

$$\sup_{\mathbf{x}\in\mathbb{R}^d}|p(\mathbf{x}) - p_a(\mathbf{x})| \leq \epsilon$$

where $\epsilon > 0$. Thus an arbitrary spectral measure $\Gamma(\mathbf{s})$ can be approximated by a discrete measure $\Gamma_a(\mathbf{s})$ such that the corresponding densities are arbitrarily close. The only requirement is that sampling of $\mathbf{s}$ is sufficiently dense.

The ch.f. corresponding to the approximated discrete spectral measure $\Gamma_a(\mathbf{s})$ sampled in L points, can be written as

$$\Phi_a(\mathbf{t}) = \exp\left(-\sum_{n=1}^{L}\psi_\alpha(\mathbf{t}^T\mathbf{s}_n)\Gamma_a(\mathbf{s}_n)\right)$$

Now define a vector $\boldsymbol{\Gamma} = [\Gamma_a(\mathbf{s}_1)\dots\Gamma_a(\mathbf{s}_L)]^T$ containing the L values of the approximated discrete spectral measure. If we evaluate the approximated ch.f. for L values of $\mathbf{t}$ and define the function $I_a(\mathbf{t}) = -\ln\Phi_a(\mathbf{t})$ then we can formulate the set of linear equations

$$\mathbf{I} = \boldsymbol{\Psi}\boldsymbol{\Gamma} \quad (6)$$

where

$$\mathbf{I} = \begin{bmatrix} I_a(\mathbf{t}_1) \\ \vdots \\ I_a(\mathbf{t}_L)\end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix}\psi_\alpha(\mathbf{t}_1^T\mathbf{s}_1) & \dots & \psi_\alpha(\mathbf{t}_1^T\mathbf{s}_L) \\ \vdots & \ddots & \vdots \\ \psi_\alpha(\mathbf{t}_L^T\mathbf{s}_1) & \dots & \psi_\alpha(\mathbf{t}_L^T\mathbf{s}_L)\end{bmatrix}$$

If the L values of $\mathbf{t}$ are chosen so that Ψ^{-1} exists, then the approximated spectral measure is given exact by the solution to (6).

3.2. Estimation of the Discrete Spectral Measure

In [6] a method for estimation of the stable spectral measure is proposed. The principle behind the estimation is based on (6). From this point, without loss of generality and to simplify the presentation, we will assume that $\mathbf{x}$ is SαS. For estimation of the spectral measure in the non-symmetric case see the more general treatment in [6].

In the case of symmetric density function the spectral measure and the ch.f. is real valued and symmetric. From the definition of the ch.f. an estimate based on samples of the random vector $\mathbf{x}$ can be obtained as

$$\hat{\Phi}(\mathbf{t}) = \frac{1}{K}\sum_{k=1}^{K}\cos(\mathbf{t}^T\mathbf{x}_k) \quad (7)$$

for K samples of $\mathbf{x}$. By introducing the function $\hat{I}(\mathbf{t}) = -\ln(\hat{\Phi}_\alpha(\mathbf{t}))$ and substitute $\mathbf{I}$ by $\hat{\mathbf{I}} = [\hat{I}(\mathbf{t}_1)\ldots\hat{I}(\mathbf{t}_L)]^T$, an estimate of the approximate discrete spectral measure is directly obtained from (6).

The spectral measure is defined on the d-dimensional unit sphere. If no a priori knowledge about the spectral measure is available and if all directions are of equal importance, it is natural to sample $\mathbf{s}$ uniformly on the unit sphere. In the SαS case it suffice to sample on the half d-dimensional unit sphere. In the $d = 2$ case the nth sample point in $\mathbf{s}$ is then $\mathbf{s}_n = [\cos\left(\pi\frac{n-1}{L}\right) \quad \sin\left(\pi\frac{n-1}{L}\right)]^T$.

The natural choice of the sampling grid of the ch.f. is to sample symmetrically on a d-dimensional sphere, again in the SαS case it suffices to sample on a half d-dimensional sphere. Thus the nth sample point in $\mathbf{t}$ is $\mathbf{t}_n = r\ \mathbf{s}_n$, where r is the sampling radius. The sampling radius is decisive for the performance of the estimator of the spectral measure, this is discussed in section 3.4.

3.3. Estimation Procedure

To avoid negative values in the estimate of the spectral measure $\mathbf{\Gamma}$, the solution in (6) is restated as

$$\min \left\| \hat{\mathbf{I}} - 2\mathrm{Re}(\mathbf{\Psi})\mathbf{\Gamma} \right\|_2 \quad \text{s.t.} \quad \mathbf{\Gamma} \geq \mathbf{0} \tag{8}$$

The estimation procedure for the spectral measure is now: determine the sampling radius of $\mathbf{t}$, calculate $\mathbf{\Psi}$, estimate $\hat{\mathbf{I}}$, and solve the constrained least square problem in (8).

3.4. Estimation Considerations

This section examine the statistical characteristics of the estimator subject to the sampling radius of the ch.f., and the choice of the characteristic exponent. To evaluate the performance of the estimation of the spectral measure we define the performance measure

$$\Upsilon = \mathrm{E}\ (\mathbf{I} - \hat{\mathbf{I}})^{\mathrm{T}}(\mathbf{I} - \hat{\mathbf{I}})/(\mathbf{I}^{\mathrm{T}}\mathbf{I})$$

where $\mathbf{I}$ is defined in (6) and $\hat{\mathbf{I}}$ is the estimate of $\mathbf{I}$ that corresponds to the estimate of the spectral measure from (8).

Applying the estimation procedure from section 3.3, we estimate the spectral measure for the SαS random vector $\mathbf{x} \in \mathbb{R}^2$ with the discrete spectral measure

$$\Gamma(\mathbf{s}_n) = \begin{cases} \frac{\gamma}{2} & \text{for } n = 11, 36\ (101, 126) \\ 0 & \text{else} \end{cases}$$

where $\mathbf{s_n}$ is distributed uniformly on the half unit circle, and $L = 90$.

A desirable property is to have a scale independent estimator, and from (2) one directly obtain that if $r \propto \gamma_x^{-1/\alpha}$ then the estimator of the ch.f. is scale independent[1].

[1] In [6] the sampling radius is fixed to $r = 1$, and the influence of the sampling radius is not addressed.

In Fig.1 the performance measure versus the sampling radius is shown for $\gamma = 0.01, 0.1, 1, 10, 100$. It appears that

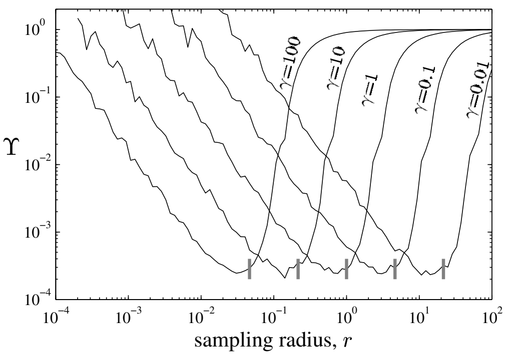


Fig. 1. Performance measure versus sampling radius, $\mathbf{x} \in$ SαS$(1.5, \gamma)$.

the sampling radius of the ch.f. is of crucial importance. A natural choice of the sampling radius is

$$r = \gamma_x^{-1/\alpha} \tag{9}$$

and the vertical bars in Fig.1 indicate the sampling radius that corresponds to (9).

In Fig.2 the performance measure versus the choice of the characteristic exponent is shown for $\mathbf{x} \in$ SαS$(\alpha, 1)$, and $\alpha = 1.2, 1.4, 1.6, 1.8$. Obviously the best performance is

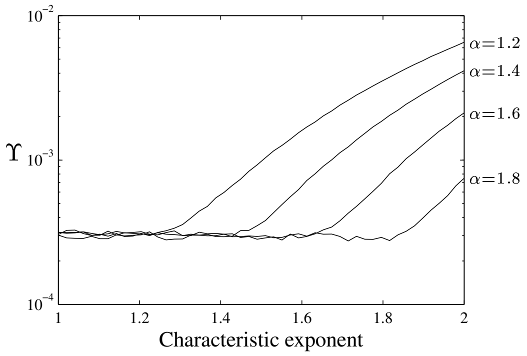


Fig. 2. Performance measure versus the choice of characteristic exponent in the estimation. $\mathbf{x} \in$ SαS$(\alpha, 1)$.

obtained when the choice of characteristic exponent in the estimation is equal to the characteristic exponent of the SαS vector. It is surprising that the characteristic exponent has only minor influence on the performance of the estimator, and notice that the characteristic exponent in the estimation should be chosen conservatively low.

4. BLIND IDENTIFICATION OF BASIS VECTORS

Consider the mixture $\mathbf{x} = \mathbf{A}\mathbf{v}$, where $\mathbf{v} = [v_1 \dots v_N]^T$ is a random vector with independent $\alpha\mathrm{S}(\alpha, \beta_n, \gamma_n, \mu_n)$ components, and $\mathbf{A} \in \mathbb{R}^{M \times N}$. The objective is to estimate the basis vectors of $\mathbf{A}$ based on observations of $\mathbf{x}$.

Fundamentally blind identification has a scale ambiguity between γ_n of v_n and the norm of the columns $\|\mathbf{a}_{\cdot n}\|_2$, thus without loss of generality we can assume that $\|\mathbf{a}_{\cdot n}\|_2 = 1$. Moreover, blind identification provides no ordering of the basis vectors, and for the case of symmetric densities there is a sign ambiguity of the basis vectors.

Now considering (5) leads to the conclusion that the identification of the basis vectors $\mathbf{a}_{\cdot n}$ is simply to determine the directions in which the spectral measure has mass[2]. In general the spectral measure has mass in $2N$ directions (N directions if we only sample the half d-dimensional sphere). The method will work for both the under-complete ($M > N$) and the over-complete ($M < N$) case.

4.1. Practical Considerations

Due to a finite number of samples, observation noise and possible deviations from the theoretical distributions, there will be some noise in the estimated spectral measure. The basis vectors should be determined as the directions in which the estimated spectral measure has dominating mass.

The number of sample points L in $\mathbf{t}$ and $\mathbf{s}$ determines the angular resolution of the basis vectors. In the $d = 2$ case with sampling on the half 2-dimensional sphere the resolution is $\pm\pi/(2L)$. A disadvantage of the proposed method is that the optimization problem in (8) scales exponentially in the dimension of the mixture.

5. SIMULATION STUDY

The basis vectors for over-complete mixtures ($M < N$) are identified for both synthetic SαS signals and speech signals.

5.1. Identification of Basis Vectors, SαS Signal

Let $\mathbf{v}$ be a random vector with independent SαS$(1.2, 1)$ random variables, $\mathbf{x} = \mathbf{A}\mathbf{v}$ be the observable random vector, and

$$\mathbf{A} = \begin{bmatrix} \cos(\theta_1) & \cos(\theta_2) & \cos(\theta_3) & \cos(\theta_4) \\ \sin(\theta_1) & \sin(\theta_2) & \sin(\theta_3) & \sin(\theta_4) \end{bmatrix}$$

with $\theta_1 = (\frac{\pi}{9})$, $\theta_2 = (\frac{2\pi}{9})$, $\theta_3 = (\frac{3\pi}{9})$, and $\theta_4 = (\frac{4\pi}{9})$.

The basis vectors are identified as the directions in which the estimated spectral measure has significant mass. In Fig.3 a typical example of an estimated spectral measure, using the estimation procedure in Sec.3, is shown. Observe that

[2]On the contrary hereto, conventional ICA algorithms is based on minimization of contrast functions.

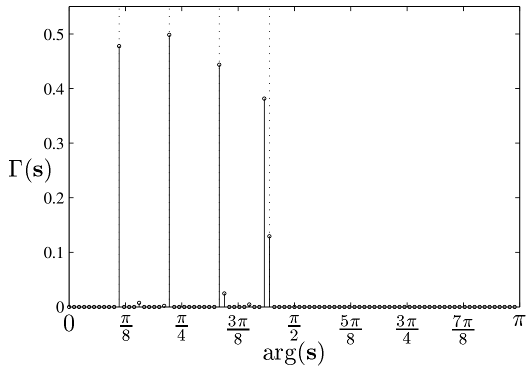


Fig. 3. Empirical spectral measure, $\Gamma(\mathbf{s})$, for the SαS random vector $\mathbf{x}$. The estimation is based on $K = 10000$ observations of $\mathbf{x}$, sampling radius $r = \frac{1}{2}$, and $L = 90$. The direction of the basis vectors of $\mathbf{A}$ are indicated by dotted lines.

the distribution of masses for the spectral measure is very specific in the 4 directions corresponding to the directions of the basis of $\mathbf{A}$. In Fig.4 the scatter plot for the observations of $\mathbf{x}$ is depicted, and the basis vectors of the mix-

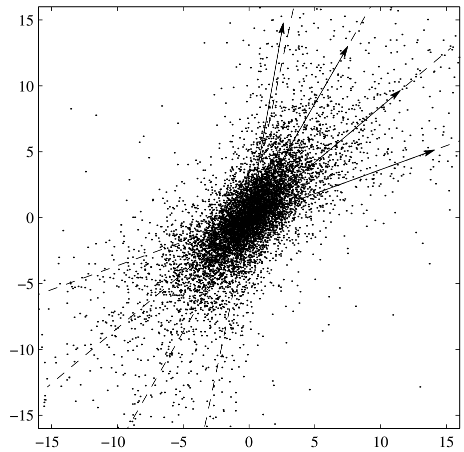


Fig. 4. Scatter plot of $\mathbf{x}$. The basis vectors of $\mathbf{A}$, are depicted by vectors, and the directions in which the spectral measure has significant masses, are depicted by dashed lines.

ing matrix and the directions in which the spectral measure has dominating mass are shown. Three basis vectors are correctly identified, and the fourth basis vector is one point away from the right value.

5.2. Identification of Basis Vectors, Speech Signals

The proposed method for identification of bases in a mixture is applied to a mixture of three speech signals. In [3] it

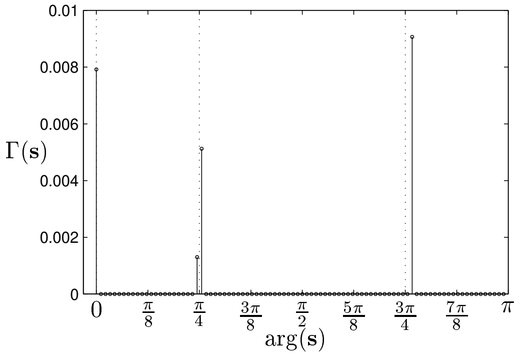


Fig. 5. Empirical spectral measure, $\Gamma(\mathbf{s})$, for the mixture $\mathbf{x}$. The ch.f. is estimated based on $N = 23000$ observations, sampling radius $r = 50$, $\alpha = 1.2$, and $L = 90$.

is demonstrated that SαS distributions are suitable for modelling a broad class of acoustical signals, including speech signals, and this is the theoretical justification for the example. For the purpose of comparison the method is applied to the same mixing matrix and the same speech signals as used in [7]. The speech signals are from the TIMIT database, and the mixing matrix is $\mathbf{A} = [1 \quad 1/\sqrt{2} \quad 1/\sqrt{2}\ ;\ 0 \quad 1/\sqrt{2} \quad -1/\sqrt{2}]$. In Fig.5 the estimated spectral measure is depicted. The distribution of masses is very specific in the directions corresponding to directions of the basis vectors of $\mathbf{A}$. In Fig.6 the scatter plot for the observations is depicted. The spectral measure has dominant masses in (2×)3 symmetric directions, and these directions are, very close to the directions of the basis vectors of $\mathbf{A}$.

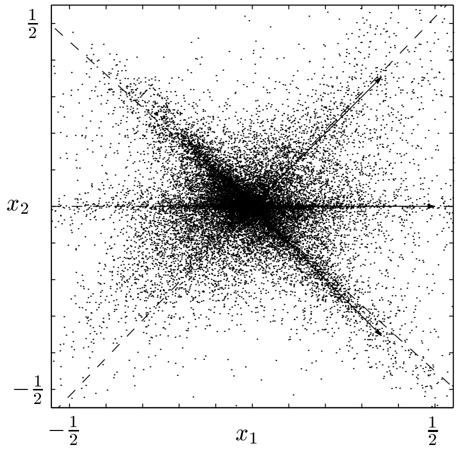


Fig. 6. Scatter plot of $\mathbf{x}$. The directions of the basis vectors of $\mathbf{A}$ are depicted by vectors, and the directions in which the spectral measure has masses, are depicted by dashed lines

6. CONCLUSION

We proposes an ICA method based on the observation that the spectral measure is discrete for stable random vectors with independent components. The method identifies the number of independent components and the non-orthogonal bases of the mixture.

We examine the properties of the estimator of the spectral measure, and discus the sampling radius for the ch.f..

Simulations on synthetic and speech signals demonstrates that the method can identify the number of independent components and the bases of the overcomplete mixtures.

Acknowledgements

Thanks to J.P.Nolan for making [6] available prior to publication, and to Lars Kai Hansen for discussions and comments on the manuscript.

7. REFERENCES

[1] Chrysostomos L. Nikias and Min Shao, *Signal Processing with Alpha-Stable Distribution and Applications*, Wiley, 1 edition, 1995.

[2] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-Stable Modeling of Noise and Robust Time-Delay Estimation in the Presence of Impulsive Noise," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 291–301, September 1999.

[3] P. Kidmose, "Alpha-Stable Distributions in Signal Processing of Audio Signals," in *41st Conference on Simulation and Modelling, Scandinavian Simulation Society*, 2000, pp. 87–94.

[4] Gennady Samorodnitsky and Murad S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman&Hall, 1994.

[5] T. Byczkowski, J. P. Nolan, and B. Rajput, "Approximation of Multidimensional Stable Densities," *Journal of Multivariate Analysis*, vol. 46, no. 1, pp. 13–31, 1993.

[6] J. P. Nolan, A. K. Panorska, and J. H. McCulloch, "Estimation of Stable Spectral Measures," *To appear in Mathematical and Computer Modelling*, 2001, Available from http://www.cas.american.edu/~jpnolan/.

[7] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations," *IEEE Signal Processing Letters*, vol. 6, pp. 87–90, April 1999.

NOMENCLATURE

The following lists defines special symbols, notation and abbriviations used in the thesis.

List of Symbols

$*$	Convolution
$.^T$	Transpose
$.^H$	Hermitian
$.^*$	Complex conjugate
$.^{\langle p \rangle}$	Signed power
$\overset{d}{=}$	Identical in distribution
$\xrightarrow{d}$	Convergence in distribution
$\sim$	Follow in distribution
$\perp$	Orthogonal
j	Imaginary unit
$\ln \cdot$	Natural logarithm
$\log \cdot$	Common base 10 logarithm
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}$	Set of real numbers
$\delta(\cdot)$	Kronecker delta function
$\Gamma(\cdot)$	Spectral measure.
$\Gamma(\cdot)$	Gamma function, $\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt$
$\phi(\cdot)$	Characteristic function
$\psi(\cdot)$	Second characteristic function

Notation

$\lvert \cdot \rvert_p$	p-norm
$\lVert \cdot \rVert_p$	Vector p-norm
$\alpha\mathrm{S}(\alpha, \beta, \gamma, \mu)$	Stable distribution with characteristic exponent α, dispersion γ, symmetry parameter β, and location parameter μ.
$\mathrm{S}\alpha\mathrm{S}(\alpha, \gamma)$	Symmetric stable distribution with characteristic exponent α and dispersion γ.
$u \sim \mathrm{S}\alpha\mathrm{S}(\alpha, \beta)$	u follow the distribution $\mathrm{S}\alpha\mathrm{S}(\alpha, \gamma)$
$\mathbf{A}$	Bold upper-case letters are used for matrices, $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{bmatrix} = [\mathbf{a}_{\cdot 1} \dots \mathbf{a}_{\cdot N}] = \left[\mathbf{a}_{1\cdot}^T \dots \mathbf{a}_{M\cdot}^T\right]^T$ $\mathbf{a}_{\cdot n}$ denote the nth column, and $\mathbf{a}_{m\cdot}^T$ denote the mth row of $\mathbf{A}$.
$\mathbf{a}$	Bold lower-case letters are used for vectors, $\mathbf{a} = [a_1 \;\; a_2 \;\; \dots \;\; a_M]^T$
$\mathbf{a}_{\cdot n}$	nth column of $\mathbf{A}$, $\mathbf{a}_{\cdot n} = [a_{1n} \;\; a_{2n} \;\; \dots \;\; a_{Mn}]^T$
$\mathbf{a}_{m\cdot}$	mth row of $\mathbf{A}$, $\mathbf{a}_{m\cdot} = [a_{m1} \;\; a_{m2} \;\; \dots \;\; a_{mN}]$
$\mathrm{E}(\cdot)$	Expectation value
$\mathrm{cum}(\cdot)$	Cumulant function.
$\mathrm{mom}(\cdot)$	Moment function.
$N(\mu, \sigma)$	Gaussian distribution, mean value μ and standard deviation σ.
$NIG(\alpha, \beta, \delta, \mu)$	Normal Inverse Gaussian distribution, α is the tail parameter, β is the symmetry parameter, δ is the scaling parameter, and μ is the location parameter.
$SG(\mathbf{R})$	Sub-Gaussian distribution.

List of Abbreviations

αS	α Stable
AR	Autoregressive
ARMA	Autoregressive Moving Averages
cgf.	Cumulant Generating Function
chf.	Characteristic Function
DCA	Dynamic Component Analysis
FLOM	Fractional Lower Order Moment
FLOS	Fractional Lower Order Statistics
g.cgf.	Generalized Cumulant Generating Function
g.chf.	Generalized Characteristic Function
HOS	Higher Order Statistics
ICA	Independent Component analysis
iid.	Independent Identical Distributed
MA	Moving Averages
mgf.	Moment Generating Function
MO	Median Orthogonal
NIG	Normal Inverse Gaussian
NOM	Negative Order Moment
NOS	Negative Order Statistics
pdf.	Probability Density Function
SαS	Symmetric α Stable

BIBLIOGRAPHY

[1] A. Ziehe and K. R. Müller. TDSEP - An Efficient Algorithm for Blind Separation Using Time Structure. In *Proceedings of the 8th International Conference on Artificial Neural Networks*, pages 675–680, 1998.

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Tables*. Dower Publications, 10 edition, 1972.

[3] Robert J. Adler, Raisa E. Feldman, and Murad S. Taqqu. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhauser, 1998.

[4] J. B. Allen and D. A. Berkley. Image Method for Efficiently Simulating Small-Room Acoustics. *Journal of Acoustical Society of America*, 65(4):943–950, April 1979.

[5] S. Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, 10:251–276, 1998.

[6] H. Attias and C. E. Schreiner. Blind Source Separation and Deconvolution: The Dynamic Component Analysis Algoritm. *Neural Computation*, 10:1373–1424, 1998.

[7] G. Aydín, O. Aríkan, and A. E. Çetin. Robust Adaptive Filtering Algorithms for α-Stable Processes. *IEEE transactions on Circuits and Systems - II*, 46(2):198–202, February 1999.

[8] O. E. Barndorff-Nielsen. Normal Inverse Gaussian Distribution and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics*, 24:1–13, 1997.

[9] O. E. Barndorff-Nielsen. Processes of Normal Inverse Gaussian Type. *Finance and Stochastics, Springer-Verlag*, 2:41–68, 1998.

[10] A. J. Bell and T. J. Sejnowski. An Information-Maximisation Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[11] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines. Blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997.

[12] J. S. Bodenschatz and C. L. Nikias. Recursive Local Orthogonal Filtering. *IEEE Transactions on Signal Processing*, 45(9):2293–2300, 1997.

[13] J. S. Bodenschatz and C. L. Nikias. Symmetric Alpha-Stable Filter Theory. *IEEE Transactions on Signal Processing*, 45(9):2301–2306, 1997.

[14] J. Borish. Extension of the Image Model to Arbitrary Polyhedra. *Journal of Acoustical Society of America*, 75(6):1827–1836, July 1984.

[15] G. E. P. Box and M. E. Muller. A Note on Generation of Random Normal Deviates. *The Annals of Mathematical Statistics*, 29:610–611, 1958.

[16] J. J. Bussgang. Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals. Technical report, Massachusetts Institute of Technology, Res. Lab. Electronics., Marts 1952. Technical Report 216.

[17] T. Byczkowski, J. P. Nolan, and B. Rajput. Approximation of Multidimensional Stable Densities. *Journal of Multivariate Analysis*, 46(1):13–31, 1993.

[18] S. Cambanis, C. D. Hardin, and A. Weron. Innovations and Wold Decompositions of Stable Sequences. *Probability Theory and Related Fields*, 79:1–27, 1988.

[19] S. Cambanis and C. Houdre'. Stable Processes: Moving Averages Versus Fourier Transforms. *Probability Theory and Related Fields*, 95:75–85, 1993.

[20] S. Cambanis and G. Miller. Linear Problems in pth Order and Stable Processes. *SIAM, Journal of Applied Mathematics*, 41(1):43–69, 1981.

[21] J. F. Cardoso. Infomax and Maximum Likelihood for Blind Source Separation. *IEEE Signal Processing Letters*, 4(4):112–114, April 1997.

[22] J. F. Cardoso. Blind Signal Separation: Statistical Principles. *Proceeding of the IEEE*, 86(10):2009–2025, 1998.

[23] J. F. Cardoso and B. H. Laheld. Equivariant Adaptive Source Separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, December 1996.

[24] J. F. Cardoso and A. Souloumiac. Blind Beamforming for Non-Gaussian Signals. *IEE Proceedings-F*, 140(6):362–370, December 1993.

[25] J. F. Cardoso and A. Souloumiac. Jacobi Angles for Simultaneous Diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164, January 1996.

[26] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[27] P. Comon. Independent component analysis: A new concept? *Signal Processing*, 36:287–314, 1994.

[28] P. Comon, C. Jutten, and J. Herault. Blind Separation of sources, Part II: Problem Statement. *Signal Processing*, 24:11–20, 1991.

[29] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, 1991.

[30] H. Cramer. *Mathematical Methods in Statistics*. Princeton University Press, 1948.

[31] S. C. Douglas. A Family of Normalized LMS Algorithms. *IEEE Signal Processing Letters*, 1(3):49–51, March 1994.

[32] S. C. Douglas. Self-Stabilized Gradient Algorithms for Blind Source Separation with Orthogonality Constraints. *IEEE Transactions on Neural Networks*, 11(6):1490–1497, November 2000.

[33] A. Edelman, T. A. Arias, and S. T. Smith. The Geometry of Algorithms with Orthogonality Constraints. *Siam, Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[34] L. Eldén and H. Park. A Procrustes Problem on the Stiefel Manifold. *Numeriche Mathematik*, 82:599–619, 1999.

[35] T. Fabricius, P. Kidmose, and L. K. Hansen. Dynamic components of linear stable mixtures from fractional low order moments. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3957–3960, 2001.

[36] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley, 2 edition, 1966.

[37] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley, 1966.

[38] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis. Alpha-Stable Modeling of Noise and Robust Time-Delay Estimation in the Presence of Impulsive Noise. *IEEE Transactions on Multimedia*, 1(3):291–301, September 1999.

[39] G. B. Giannakis. Cumulants: A Powerful Tool in Signal Processing. *Proceedings of the IEEE*, 75(9):1333–1334, September 1987.

[40] M. Girolami. An Alternative Perspective on Adaptive Independent Component Analysis Algorithms. *Neural Computation*, 10:2103–2114, 1998.

[41] M. Girolami. *Advances in Independent Component Analysis.* Springer-Verlag, New York, 2000.

[42] M. Girolami and C. Fyfe. Stochastic ICA Contrast Maximisation Using Oja's Nonlinear PCA Algorithm. *International Journal of Neural Systems*, 8:661–678, 1997.

[43] L. K. Hansen and J. Larsen. Source separation in short image sequences using delayed correlation. *Proceedings of the IEEE Nordic Signal Processing Symposium, Vigsø, Denmark 1998. Eds. P. Dalsgaard and S.H. Jensen*, pages 253–256, 1998.

[44] L. K. Hansen, J. Larsen, and T. Kolenda. *"Multimedia Image and Video Processing, Eds.: L. Guan, S. Y. Kung and J. Larsen*, chapter 7, pages 175–199. CRC Press, 2000. On Independent Component Analysis for Multimedia Signals.

[45] P. S. K. Hansen and J. A. Sørensen. Real Time Multi-Modal Communication: Audio Interface. Version 0.2. Unpublished. This report is to be published as a Technical Report from Informatics and Mathematical Modelling, Technical University of Denmark., August 1999.

[46] A. Hanssen and T. A. Øigård. The Normal Inverse Gaussian Distribution as a Flexible Model for Heavy-Tailed Processes. In *Proceeding of NSIP*, 2001.

[47] Simon Haykin. *Adaptive Filter Theory*. Prentice-Hall, 3 edition, 1996.

[48] Russel H.Lambert. *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath mixtures*. PhD thesis, University of Southern California, 1996.

[49] C. L. Isbell and P. Viola. Restructuring Sparse High Dimensional Data for Effective Retrieval. *Advances in Neural Information Processing Systems*, 11:480–486, 1998.

[50] Don H. Johnson and Dan E. Dudgeon. *Array Signal Processing: Concepts And Techniques*. Prentice-Hall, 1 edition, 1993.

[51] C. Jutten and J. Herault. Blind Separation of Sources, Part I: An Adaptive Algortihm based on Neuromimetic Architecture. *Elsevier, Signal Processing*, 24:1–10, 1991.

[52] Thomas Kailath. *Linear Systems*. Prentice-Hall, 1 edition, 1980.

[53] M. Kanter and W. L. Steiger. Regression and Autoregression with Infinite Variance. *Advances in Applied Probability*, 6(4):768–783, 1974.

[54] P. Kidmose. Adaptive Filtering for Non-Gaussian Processes. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pages 424–427, 2000.

[55] P. Kidmose. Alpha-Stable Distributions in Signal Processing of Audio Signals. In *41st Conference on Simulation and Modelling, Scandinavian Simulation Society*, pages 87–94, 2000.

[56] P. Kidmose. Independent Components Analysis Using the Spectral Measure for Alpha-Stable Distributions. In *Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, pages 1–5, 2001.

[57] Preben Kidmose and Jakob Bøgelund Pedersen. Beamforming til støjreduktion i høreapparater. Master Thesis (in Danish), Department of Mathematical Modelling, Technical University of Denmark, 1998.

[58] T. Kolenda, L. K. Hansen, and S. Sigurdsson. *Advances in Independent Component Analysis, Editor: M. Girolami*, chapter 13, Independent Components in Text, pages 229–250. Springer-Verlag, 2000.

[59] J.-H. Lee, H.-Y. Jung, T.-W. Lee, and S.-Y. Lee. Speech Feature Extraction using Independent Component Analysis. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pages 1631–1634, 2001.

[60] T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski. A Unifying Information-Theoretic Framework for Independent Component Analysis. *Elsevier, Computers and Mathematics with Applications*, 39(11):1–21, 2000.

[61] T.-W. Lee and A. Ziehe. Combining Time-Delayed Decorrelation and ICA: Towards Solving the Cocktail Party Problem. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pages 1249–1252, 1998.

[62] Te-Won Lee. *Independent Component analysis: Theory and Applications*. Kluwer Academic Publishers, Boston, 1 edition, 1998.

[63] V. P. Leonov and A. N. Shiryaev. On a Method of Calculation of Semi-Invariants. *Theory of Probability and its Applications*, 4:319–329, 1959. English translation from Teorija verojatnostej i ee primenenija.

[64] Ulf A. Lindgren. *Signal separation - II.* PhD thesis, Chalmers University of Technology, 1997.

[65] Eugene Lukacs. *Characteristic Functions*. Charles Griffin and Company Limited, 1st edition, 1960.

[66] X. Ma and C. L. Nikias. Parameter Estimation and Blind Channel Identification in Impulsive Signal Environments. *IEEE Transactions on Signal Processing*, 43(12):2884–2897, December 1995.

[67] G. Miller. Properties of Certain Symmetric Stable Distributions. *Journal of Multivariate Analysis*, 8:346–360, 1978.

[68] M. J. McKeown MJ, T. P. Jung, S. Makeig, G. Brown G andS. S. Kindermann, T.W. Lee, and T. J. Sejnowski. Spatially Independent Activity Patterns in Functional MRI Data During the Stroop Color-naming Task. *Proceedings of the National Academy of Sciences of the United States of Amerika.*, 95:803–810, February 1998.

[69] L. Molgedey and H. Schuster. Separation of independent signals using time-delayed correlations. *Physical Review Letters*, 72(23):3634–3637, 1994.

[70] S. T. Neely and J. B. Allen. Invertibility of a Romm Impulse Response. *Journal of Acoustical Society of America*, 66(1):165–169, July 1979.

[71] C. L. Nikias and J. M. Mendel. Signal Processing with Higher-Order Spectra. *IEEE Signal Processing Magazine*, pages 10–37, July 1993.

[72] Chrysostomos L. Nikias and Athina P. Petropulu. *Higher-order Spectra Analysis, A Nonlinear Signal Processing Framework*. Prentice-Hall, 1 edition, 1993.

[73] Chrysostomos L. Nikias and Min Shao. *Signal Processing with Alpha-Stable Distribution and Applications*. Wiley, 1 edition, 1995.

[74] J. P. Nolan, A. K. Panorska, and J. H. McCulloch. Estimation of Stable Spectral Measures. *Mathematical and Computer Modelling*, 34:1113–1122, 2001. Available from http://www.cas.american.edu/˜jpnolan/.

[75] J. P. Nolan and B. Rajput. Calculation of Multidimensional Stable Densities. *Communications in Statistics - Simulation and Computation*, 24:551–566, 1995.

[76] John P. Nolan. Stable Distributions, Models for Heavy Tailed Data. Chapter 1, Basic Properties of Univariate Stable Distributions, August 2000.

[77] Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing*. Prentice-Hall, 1st edition, 1975.

[78] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3 edition, 1991.

[79] L. Potamitis, N. Fakotakis, and G. Kokkinakis. Independent Component Analysis Applied to Feature Extraction for Robust Automatic Speech Recognition. *IEEE Electronics Letters*, 36(23):1977–1978, November 2000.

[80] Murray Rosenblatt. *Stationary Sequences and Random Fields*. Birkhauser, 1 edition, 1985.

[81] S. T. Roweis. One Microphone Source Separation. *Advances in Neural Information Processing Systems*, 13:793–799, 2000. The MIT Press.

[82] T. H. Rydberg. The Normal Inverse Gaussian Lévy Process: Simulation and Approximation. *Communications in Statistics - Stochastic Models*, 13(4):887–910, 1997.

[83] Henrik Sahlin. *Blind Signal Separation by Second Order Statistics*. PhD thesis, Chalmers University of Technology, 1998.

[84] Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman&Hall, 1994.

[85] Daniel Schobben. *Efficient Adaptive Multi-channel Concepts in Acoustics*. PhD thesis, Technische Universiteit Eindhoven, 1999.

[86] A. Swami and B. Sadler. Parameter Estimation for Linear Alpha-Stable Processes. *IEEE Signal Processing Letter*, 5(2):48–50, February 1998.

[87] P. Switzer and A. A. Green. Min/MAX Autocorrelation Factors for Multivariate Spatial Imagery. Technical report, Department of Statistics, Stanford University, April 1984. Technical Report No. 6.

[88] L. Tong, Y. Inouye, and R. W. Liu. Waveform-Preserving Blind Estimation of Multiple Independent Sources. *IEEE Transactions on Signal Processing*, 41(7):2461–2469, July 1993.

[89] G. A. Tsihrintzis and C. L. Nikias. Fast Estimation of the Parameters of Alpha-Stable Impulsive Interference. *IEEE Transactions on Signal Processing*, 44(6):1492–1503, June 1996.

[90] A. J. Veen. Joint Diagonalization via Subspace Techniques. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, page ., 2001.

[91] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent Component Analysis for Identification of Artifacts in Magnetoencephalographic Recordings. *Advances in Neural Information Processing Systems*, 10:229–235, 1998. MIT Press.

[92] R. Vigário, Jaakko Särelš, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent Component Approach to the Analysis of EEG and MEG Recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, May 2000.

[93] A. Yeredor. Approximate Joint Diagonalization using Non-Othogonal Matrices. In *International Workshop on Independent Component Analysis and Blind Signal Separation*, page 6, 2000.

[94] A. Yeredor. Blind source Separation via the Second Characteristic Function. *Signal Processing*, 80(5):897–902, 2000.

[95] V. M. Zolotarev. *One Dimensional Stable Distributions*, volume 65 of *Translation of Mathematical Monographs*. American Mathematical Society, 1986. Translation from the Original 1983 Russian Edition.

INDEX