Helsinki University of Technology Signal Processing Laboratory
Teknillinen korkeakoulu Signaalinkäsittelytekniikan laboratorio
Espoo 2002 Report 36

ADAPTIVE METHODS FOR BLIND EQUALIZATION AND SIGNAL SEPARATION IN MIMO SYSTEMS

Mihai Enescu

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S1 at Helsinki University of Technology (Espoo, Finland) on the 23rd of August, 2002, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Signal Processing Laboratory

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Signaalinkäsittelytekniikan laboratorio

Distribution:
Helsinki University of Technology
Signal Processing Laboratory
P.O. Box 3000
FIN-02015 HUT
Tel. +358-9-451 2486
Fax. +358-9-460 224
E-mail: anne.jaaskelainen@hut.fi

ISBN 951-22-6042-5
ISSN 1456-6907

Otamedia Oy
Espoo 2002

Abstract

This thesis addresses the problems of blind source separation (BSS) and blind and semi-blind communications channel equalization. In blind source separation, signals from multiple sources arrive simultaneously at a sensor array, so that each sensor output contains a mixture of source signals. Sets of sensor outputs are processed to recover the source signals from the mixed observations. The term *blind* refers to the fact that specific source signal values and accurate parameter values of a mixing model are not known *a priori*. Application domains for the material in this thesis include communications, biomedical, and sensor array signal processing.

The goal of this thesis is development of blind and semi-blind algorithms which require little or no prior information about source signal or mixing system parameter values in order to process the data. We start with the problem of extracting unknown input signals from measured outputs of instantaneous multiple-input multiple-output (I-MIMO) systems with constant parameter values. Suggested solutions are then extended to time-varying I-MIMO systems and also to constant finite impulse response multiple-input multiple-output (FIR-MIMO) systems. Another goal is to find a practical solution for the more challenging case of time-varying FIR-MIMO systems.

The source separation techniques proposed in this thesis are based on state-space models and on recursive estimation. Blind separation algorithms based on Kalman filters are proposed. The source signals are treated using low-order autoregressive models. Projections along signal subspace eigenvectors are used to reduce the dimensionality of observations and also for spatial decorrelation of sources. Any changes that occur in the signal subspace can be tracked online. When considering slowly time-varying FIR-MIMO systems, fractional sampling can be used to derive a set of slowly time-varying I-MIMO systems. Thus, the proposed recursive BSS algorithms for I-MIMO systems can be used for blind equalization of slowly time-varying FIR communications channels.

The problem of equalization of time-varying FIR MIMO systems is also addressed in this thesis. The proposed solutions involve semi-blind algorithms which work in two stages. First,

a channel estimate is derived, and then the observation sequence is equalized. The algorithms estimate the otherwise-unknown noise statistics, and as a result achieve performance close to that of an optimal Kalman-based algorithm. A non-connected decision feedback equalization algorithm is derived for FIR-MIMO systems, using a minimum mean square error criterion. Simulation results show that the algorithm is able to track time and frequency selective channels and also to mitigate intersymbol and interuser interference.

Preface

The research work for this thesis was carried out at the Signal Processing Laboratory, Helsinki University of Technology and the Signal Processing Laboratory, Tampere University of Technology, during the years 1999-2001. I wish to express my gratitude to my supervisor, Prof. Visa Koivunen, for his encouragement, guidance and support during the course of this work.

I would like to thank my thesis pre-examiners Prof. Ioan Tabus and Dr. Juha Laurila for their constructive comments. I am also grateful to Prof. Iiro Hartimo, the director of Graduate School in Electronics Telecommunications and Automation (GETA), to GETA secretary Marja Leppäharju, and to our laboratory secretaries Anne Jääskeläinen and Mirja Lemetyinen for helping me with many practical issues and arrangements.

Many thanks to Dr. Charles Murphy for revising the language and for the great moments we had in the U.S. and in Finland. I would like to thank Marius Sirbu, who co-authored several of my publications. I am also grateful to my colleagues and co-workers, Dr. Yinglu Zhang, Dr. Samuli Visuri, Jan Eriksson, Dr. Jukka Mannerkoski, Dr. Doru Giurcaneanu, Maarit Melvasalo, Juha Karvanen, Timo Roman, Traian Abrudan, and Stefan Werner for our many interesting discussions. Thanks are also due my friends in Helsinki and in Tampere with whom I spent wonderful moments.

The financial support of the Academy of Finland, GETA, Nokia Foundation, Finnish Society of Electronics Engineers Foundation, Tekniikan Edistämissäätiö, Emil Aaltosen Säätiö are gratefully acknowledged.

I wish to give my deepest gratitude to my family, especially to my mother for her permanent encouragement and support. Finally, I would like to thank my lovely fiancée Monica for love and understanding during my time as a graduate student.

Otaniemi, June 2002

Mihai Enescu

Contents

List of publications

I

M. Enescu, V. Koivunen. Tracking Time-Varying Mixing System in Blind Separation. In *IEEE Workshop on Sensor Array and Multichannel (SAM) Signal Processing*, pp. 291–295, March 2000.

II

M. Enescu, V. Koivunen. Recursive Estimator for Separation of Arbitrarily Kurtotic Sources. *IEEE Workshop on Statistical Signal and Array Processing (SSAP)*, pp. 301–305, August 2000.

III

M. Enescu, Y. Zhang, S.A. Kassam, V. Koivunen. Recursive Estimator for Blind MIMO Equalization via BSS and Fractional Sampling. In *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 94–97, March 2001.

IV

V. Koivunen, M. Enescu, E. Oja. Adaptive Algorithm for Blind Separation from Noisy Time-Varying Mixtures. In *Neural Computation*, 13 (10): pp. 2339–2357, October 2001.

V

M. Enescu, M. Sirbu, and V. Koivunen. Recursive Semi-Blind Equalizer for Time-Varying MIMO Channels. In *IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 289–292, August 2001.

VI

M. Enescu, M. Sirbu, V. Koivunen. Adaptive Equalization of Time-Varying MIMO Channels. *Report 34, Signal Processing Laboratory, Helsinki University of Technology, ISBN 951-22-5944-3, submitted to Signal Processing*, May 2001.

VII

M. Enescu, M. Sirbu, V. Koivunen. Recursive Estimation of Noise Statistics in Kalman Filter Based MIMO Equalization. In press, *URSI General Assembly*, August 2002.

List of abbreviations and symbols

Abbreviations

AR	Autoregressive
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase Shift Keying
BSS	Blind Source Separation
BU	Bad Urban
COST	European cooperation in the field of Scientific and Technical research
CDMA	Code Division Multiple Access
DFE	Decision Feedback Equalizer
DQPSK	Differential Quadrature Phase Shift Keying
EASI	Equivariant Adaptive Separation via Independence
ECG	Electrocardiogram
EEG	Electroencephalogram
FB	Feedback
FC	Fully-Connected
FF	Feedforward
FIR	Finite Impulse Response
GSM	Global System for Mobile Communications
HOS	Higher Order Statistics
HT	Hilly Terrain
I	Instantaneous
ICA	Independent Component Analysis
ICASSP	International Conference on Acoustics Speech and Signal Processing
i.i.d.	Independent, Identically distributed
ISI	Intersymbol Interference
IUI	Interuser Interference
JADE	Joint Approximate Diagonalization of Eigen-matrices
LMS	Least Mean Squares
LOS	Line Of Sight
LTI	Linear Time Invariant
MBD	Multichannel Blind Deconvolution
MDL	Minimum Description Length
MEG	Magnetoencephalogram
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MLSE	Maximum Likelihood Sequence Estimator

MSE	Mean Square Error
M-PSK	M-ary Phase Shift Keying
NC	Non-Connected
OFDM	Orthogonal Frequency Division
PAST	Projection Approximation Subspace Tracking
PASTd	Projection Approximation Subspace Tracking with deflation
PCA	Principal Component Analysis
pdf	probability density function
RA	Rural Area
RLS	Recursive Least Squares
SER	Symbol Error Rate
SIMO	Single-Input Multiple-Output
SIS	Sequential Importance Sampling
SISO	Single-Input Single-Output
SHIBBS	Shifted Blocks for Blind Separation
SNR	Signal to Noise Ratio
SVD	Singular Value Decomposition
TDMA	Time Division Multiple Access
TS	Training Sequence
TU	Typical Urban
TV	Time-varying
TVC	Time-varying Channel
WSSUS	Wide Sense Stationary with Uncorrelated Scattering

Symbols

a^*	complex conjugate of the scalar a
$\mathbf{a}^H$	Hermitian transpose of the vector $\mathbf{a}$
$\bar{\mathbf{a}}$	a vector which contains other vectors $\mathbf{a}_i$
$\tilde{\mathbf{a}}$	mean of a vector $\mathbf{a}$
$\hat{\mathbf{a}}$	estimate of a vector $\mathbf{a}$
$\mathbf{A}^T$	transpose of the matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of the matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of the matrix $\mathbf{A}$
$\bar{\mathbf{A}}$	a matrix which contains other matrices
i, j, l, k	indices
t	continuous time variable
N	number of samples
N_p	number of particles
N_l	number of echo paths
T	symbol period
M	oversampling factor
m	number of sources/transmitters
n	number of sensors/receivers
$p(\mathbf{s})$, $q(\mathbf{s})$	probability density function of $\mathbf{s}$
$\mathbf{s}_k$	source vector $\mathbf{s}$ at time k
s_i	ith source s
$\mathbf{y}_k$	observation vector $\mathbf{y}$ at time k
$\mathbf{x}_k$	separated vector of sources at time k in BSS
$\mathbf{z}_k$	whitened observations vector in BSS
$x_i(k)$	kth transmitted symbol by user i in communications
$\hat{x}$	estimated symbol in communications
z_k	soft decision in DFE
$\mathbf{A}$	mixing matrix in BSS

$\mathbf{W}$	separation matrix in BSS
$\mathbf{W}^o$	orthogonal separation matrix in BSS
$\mathbf{W}_k^l$	lth separation matrix at time k in blind deconvolution
$\mathbf{V}$	whitening matrix
$\mathbf{F}$	state transition matrix in state-space models
$\mathbf{K}$	Kalman gain matrix
$\mathbf{k}$	gain vector
$\mathbf{H}$	MIMO channel matrix
$\mathbf{P}(k\|k-1)$	prediction error covariance matrix
$\mathbf{P}(k\|k)$	filtering error covariance matrix
$\boldsymbol{\mathcal{P}}$	permutation matrix
$\boldsymbol{\mathcal{S}}$	scaling matrix
$\mathcal{K}$	Kullback-Leibler divergence
$\mathbf{I}$	identity matrix
ϕ	contrast function
ϕ^o	orthogonal contrast function
$H(x_i)$	entropies of the entries of x_i
$\mathbf{f}$	feedforward filter in DFE
$\mathbf{d}$	feedback filter in DFE
$f(\cdot), g(\cdot)$	non-linear functions
L_r	filter length for blind deconvolution
N_f	feedforward filter length
N_d	feedback filter length
$\mathbf{R}_{\mathbf{x}}$	covariance matrix of $\mathbf{x}$
$\mathbf{R}_{xx}$	auto-correlation matrix
$\mathbf{R}_{xy}$	cross-correlation matrix
C	cumulants
$\mathcal{C}$	matrix of correlation coefficients
κ_i	kurtosis of the received ith source

$\varphi(\cdot)$	score function
$\mathbf{v}_k$	observation noise vector at time k
$\mathbf{w}_k$	state noise vector at time k
r	AR model order
L	length of the channel
$\mathbf{E}$	matrix of eigenvectors
$\mathbf{D}$	diagonal matrix of eigenvalues
λ_i	ith eigenvalue
$\mathbf{e}_i$	ith eigenvector
δ	Kronecker delta
Δ	delay used in decision feedback equalizer design
τ	delay
$\mu, \alpha, \beta, \gamma$	adaptation step size parameters
h_{ij}	channel coefficient from transmitter i to receiver j
σ_v^2	observation noise variance
$\mathrm{E}[\cdot]$	expectation operator
$\|\| \cdot \|\|$	Euclidean norm
$\|\| \cdot \|\|_F$	Frobenius norm

Chapter 1

Introduction

1.1 Motivation

In blind source separation (BSS), multiple observations acquired by an array of sensors are processed in order to recover the initial multiple source signals. The term *blind* refers to the fact that there is no explicit information about the mixing process or about source signals. The concept of blind source separation is related to independent component analysis (ICA). However, ICA can be viewed as a general-purpose tool taking the place of principal component analysis (PCA) which means it is applicable to a wide range of problems. Some application domains of blind source separation are biomedical signal analysis, geophysical data processing, data mining, wireless communications and sensor array processing.

Blind source separation techniques can be traced back to the work of Herault and Jutten [59] in 1983 on a real-time algorithm used to solve the blind separation problem. In the related area of blind channel equalization, Sato (1975) [115] and Godard (1980) [52] introduced techniques for channel equalization using symbol statistics rather than known training symbol sequences. In the years following the publication of these early works, the theory and practice of blind source separation have evolved tremendously. The instantaneous multiple-input multiple-output (I-MIMO) noise-free linear model has been extended to linear FIR models and to nonlinear instantaneous models. Many different algorithms have been proposed for BSS [1, 2]. These algorithms have proven practical in varied areas of application. For instance, independent component analysis has

been used for separating contributions from different neural currents in the brain which appear as mixed observations from an EEG array. There are, of course, many other applications. A search of the IEEE publication database or other sources will reveal several thousand works related to blind methods.

In recent years, the communications community has recognized the importance of blind signal processing techniques. This is partly due to the fact that wireless communications field experienced explosive growth, and demand for high data-rate services has been increasing. Blind methods in communications use slightly different models from those used in blind source separation. For example, the distortion caused by multipath propagation in a communications channel spreads the signal in time and causes frequency selective fading. The models used in communications are convolutive rather than instantaneous. Hence, for communications, the finite impulse response multiple-input multiple-output (FIR-MIMO) model is appropriate. The term *blind* has become quite popular for describing any estimation problem in which there is fairly limited *a priori* system information.

A good reason for investigating blind techniques in the context of communications is that spectrum is a limited resource. Improved spectral efficiency and higher effective data rates are important design goals of future communication systems. Use of multiple antennas at receivers and/or at transmitters justifies MIMO models. Hence, blind techniques based on MIMO models are very practical. Conventional techniques for receiver mitigation of communications channel distortions require either knowledge of the channel parameter values or a sequence of known training symbols. In particular, channel estimation and equalization rely on training signals. This obviously decreases the effective data rate [126]. For time-invariant channels, the loss is insignificant because only one training cycle is necessary. For time-varying channels, the training has to be performed periodically, which significantly lowers the throughput. For example, in GSM, about 20% of the symbols are used for training.

Most algorithms in communications systems are batch or block oriented and assume burst transmission. Even if the channel is considered time-varying, during the burst period it is assumed to be invariant. One limitation of batch blind equalization algorithms is their nonrecursive structure, which effectively limits their applicability in time-varying scenarios requiring real-

time computation. In a slowly time-varying fading environment, blind algorithms could be used to perform equalization. However, in the case of a deep fade, during which the equalizer may lose track of the time-varying channel, batch algorithms may suffer. Structures that recursively compute new symbol estimates by considering past channel and symbol estimates are better suited for such time-varying channels [105].

Blind methods in communications are particularly appealing because they may allow all of the symbol periods allocated for sending training symbols to be used for sending data symbols instead. However, blind methods in communications have shortcomings. They may rely on unrealistic assumptions and they may also have poor convergence properties. Moreover, ambiguities always remain when blind methods are used. Hence, so called semi-blind methods provide an interesting alternative to both blind and non-blind methods. Optimal semi-blind techniques exploit the same information as blind methods, and also use the information coming from the design of the receiver [35]. By incorporating some known information, semi-blind techniques avoid the problems encountered by blind methods. They may also allow shorter training sequences, so an increase in the effective data rate can be obtained even if training sequences are not eliminated entirely. The trade-off between blind and non-blind techniques makes the semi-blind methods appealing for cost-effective and practical implementation in future receivers.

1.2 Scope of the thesis

The scope of this thesis is consideration of the problems of blind source separation, blind channel equalization, and semi-blind channel equalization. The goal of the thesis is to define complete algorithms capable of providing desired separation and equalization properties. The algorithms should use as little prior information as possible for processing observations, while achieving substantial performance improvements over existing techniques.

The main application area of the proposed algorithms is wireless communications. The design goal of the techniques is robust performance in noisy time-varying environments. Real-time computation is an important issue in the face of time-varying systems. The computational complexity of the resulting algorithms should be relatively low when compared to existing algo-

rithms. Simulation studies illustrating the performance of the algorithms should be performed in a realistic manner.

1.3 Contributions of the thesis

The contributions of this thesis are in the area of blind source separation for I-MIMO and FIR-MIMO models. Also the semi-blind FIR-MIMO equalization problem is considered in the case of time-varying channels. The main application area considered is wireless communications. However, simulations related to biomedical signal processing are also carried out.

The problem of on-line blind separation in the case of an instantaneous and slowly time-varying linear mixing system is considered first. An algorithm is proposed based on a state-space model. It employs subspace tracking and recursive estimation stemming from the Kalman filter. It is demonstrated that separation of sources and noise attenuation can be performed simultaneously. Source signals are modeled using low order autoregressive models and noise is attenuated by trading off between the model and the information provided by measurements. By using a Kalman based source separation algorithm, the observation noise is taken into account. The problems of detecting and adapting to changes that may occur in the mixing system are also addressed. Fractional sampling may be used to convert a FIR-MIMO model into a I-MIMO one [143]. Using this technique it is shown that recursive BSS can be applied to equalization of slowly time-varying channels. The performance of the separation algorithm is investigated in simulations using biomedical and communications signals at different noise levels and using a time-varying mixing system.

Recursive estimation is employed for tracking time-varying parameter values of communications channels. A semi-blind algorithm is proposed, with a short training sequence used at the beginning of transmission to acquire the statistical information needed by the Kalman-based channel estimation algorithm and also estimate the channel. After the training period ends the algorithm relies on the decisions of an equalizer, and hence operates in a decision-directed mode. The algorithm operates in two stages. In the first stage the channel is estimated and in the second stage equalization is performed based on the channel estimates. Batch and on-line methods for

estimating the unknown noise statistics needed by Kalman filter are introduced. By including a noise statistics estimation stage, less prior information is needed and improved performance is achieved because critical parameter values are estimated rather than assumed. A multiple-input multiple-output minimum mean square error decision feedback equalizer (MIMO MMSE-DFE) is also derived. Simulations are carried out based on a realistic channel model [100].

The remainder of this thesis is organized as follows. Chapter 2 introduces the signal model and the basic concepts employed in blind source separation. A brief review of the main classes of algorithms is given and several applications are described. Chapter 3 contains a review of adaptive whitening techniques. These methods are based on adaptive update of signal and noise subspaces. The problem of tracking changes in the signal subspace is also considered.

In chapter 4, an adaptive blind source separation method is introduced. A review of adaptive algorithms based on state-space model is given as well. The problem of modeling the sources or the mixing matrix is also considered.

Chapter 5 deals with adaptive algorithms for semi-blind equalization. The chapter begins with a brief introduction of the channel model that is used. The chapter focuses on algorithms which perform joint channel estimation and symbol equalization. A description of the application of Kalman filter to channel estimation and tracking is given and different derivations of the decision feedback equalizer are presented. Finally, chapter 6 summarizes the results and contribution of the thesis.

1.4 Summary of publications

The material in this thesis has appeared in seven other publications. Four relate specifically to blind source separation and three relate specifically to semi-blind equalization.

In paper *I*, the problem of blind separation of signals in time-varying mixtures is addressed. The proposed solution uses an adaptive whitening transform. A technique employing subspace tracking is proposed. A Kalman filter based algorithm is used to perform recursive blind source separation. The state transition matrix is augmented to contain a low-order autoregressive model so as to have a more accurate prediction. Tracking changes in the signal subspace is another aim

of the paper. Examples using time-varying mixtures where the signal subspace changes in time are presented with both test signals and communication signals.

Paper II is an extension of the algorithm presented in paper *I* to solve the problem of separating both sub- and super-Gaussian densities. A fully adaptive algorithm is obtained by employing a criterion for choosing a suitable zero-memory nonlinearity for each channel. In the simulation examples, electrocardiogram (ECG) signals are employed in order to demonstrate a practical application were signals with different kurtosis have to be separated and classical methods employing fixed nonlinearities for all channels fail. A slow variation of the elements of the mixing matrix is also considered.

Paper III, shows the applicability of the recursive separation algorithm to the problem of blind equalization. Based on a fractional sampling technique [143], the blind equalization problem is converted in a blind source separation (BSS)problem. Communication signals are considered in simulations and a slowly time-varying model is used for the mixing system.

Paper IV is one of the main publications of this thesis. A complete recursive algorithm for blind source separation is presented. Simulation results are reported using both medical and communications signals in different scenarios. Changes in the signal subspace are considered and slowly time-varying mixing matrix is tracked.

Paper V deals with adaptive semi-blind equalization. The problem of multiple-input multiple-output (MIMO) systems with application to communications is addressed. The time and frequency selective nature of the channels is considered. A channel model based on measurements is used in simulations. A state-space model is used to describe the system. The channel taps are stacked in a state vector. A Kalman filter is employed to estimate and track the channel. A minimum mean square error decision feedback equalizer for a system with two inputs and two outputs is derived. The joint channel estimation/symbol equalization algorithm uses a training sequence for initial parameter acquisition after that it runs in decision-directed mode. Results presenting the mitigation of both the intersymbol interference (ISI) and inter-user interference (IUI) are reported.

Paper VI is another main publication of this thesis. Adaptive equalization of time-varying MIMO channels is addressed. The results from paper V are extended to a general case. A

comprehensive derivation of MIMO minimum mean square error - decision feedback equalizer (MMSE-DFE) is presented. Another goal is to design a channel estimator which does not require too much information about the system other than knowledge of the training sequence. This means that when using the state-space approach the state and measurement noise covariances are estimated from the data. Both batch and recursive methods for estimating the noise covariances are derived in the paper. Simulation results showing estimation of noise statistics, channel tracking, and mitigation of ISI and IUI are reported.

Paper VII addresses the problem of noise estimation in Kalman filter based MIMO equalization. Estimating noise statistics is of great interest when using state-space models. Kalman filtering requires accurate values of state and measurement noise covariances to work optimally. In this paper a recursive method for estimating the noise statistics with application to equalization of time-varying MIMO channels is proposed. The optimality of the estimates is tested using non-parametric runs test on innovation sequences. The accurate estimation of noise covariance matrices allows the Kalman filter to reliably estimate the state, thus leading to improved equalization performance.

All of the simulation software for the all of the original papers of this dissertation was written solely by the author, with the exception of that used for papers IV-VII, which had contributions from the other authors. The original Kalman-filter-based separation algorithm which appeared in paper IV was the idea of the first author. The author of this thesis contributed material relating to subspace tracking, detection of changes in mixing system and selection of appropriate nonlinearities leading to a more complex recursive algorithm which is described in papers I-IV. He was mainly responsible of planning experiments for all the papers. The author derived the analytical results in paper VI and did most of the writing of papers I-III, and V-VII. The co-authors collaborated in experiment design, provided guidance for the author's proofs, and contributed to the writing of the final version of each paper.

Chapter 2

BSS model

There is plenty of recent work on blind source separation (BSS) in the signal processing, communications and neural network research communities. Recent publications include article collections [44, 56], special magazine issues [1, 2], monographs [51, 86] and two comprehensive books [28, 65]. Some review articles do exist [79] and also many articles have been published [16, 17, 20, 30, 68, 71, 75, 103].

This chapter introduces different models used in blind separation and the underlying assumptions that justify their use. Several categories of algorithms are presented and key concepts are described. The problem of blind deconvolution and a few important applications of blind separation are also discussed in brief.

2.1 Problem formulation and assumptions

The goal of blind source separation is to recover original source signals from sensor observations that are mixtures of the original source signals. Over the years, several models [31] of the mixing process have been used. We start from the basic linear model that relates the unobservable source signals and the observed mixtures:

$$\mathbf{y}_k = \mathbf{A}\mathbf{s}_k, \tag{2.1}$$

where $\mathbf{A}$ is an $n \times m$ matrix of unknown mixing coefficients, $n \geq m$, $\mathbf{s}$ is a column vector of m source signals, $\mathbf{y}$ is a column vector of n mixtures, and k is the time index. This model is

instantaneous because the mixing matrix contains fixed elements, and also *noise-free*. If noise is included in the model, it can be treated as an additional source signal or as measurement noise. In the case when it is present as measurement noise, the model becomes:

$$\mathbf{y}_k = \mathbf{A}\mathbf{s}_k + \mathbf{v}_k, \tag{2.2}$$

where the noise vector $\mathbf{v}_k$ is of dimension $n \times 1$. The mixing matrix may be constant, or can vary with the time index k. In the time-varying case, $\mathbf{A}$ becomes $\mathbf{A}_k$. In multichannel blind deconvolution or blind equalization, the n-dimensional vector of received signals $\mathbf{y}_k$ is assumed to be produced from the m-dimensional vector of source signals using the z-domain mixture model:

$$\mathbf{y}(z) = \mathbf{A}(z)\mathbf{s}(z). \tag{2.3}$$

In practice, the mixing matrix $\mathbf{A}$ contains FIR filters. Assuming that the direct paths from source i to sensor j have equal length L (see Figure 2.1), we have:

$$\mathbf{y}_k = \sum_{l=0}^{L-1} \mathbf{A}^l \mathbf{s}_{k-l} + \mathbf{v}_k. \tag{2.4}$$

Finally, if the mixing matrix is allowed to be time-varying, we will use the notation $\mathbf{A}^l = \mathbf{A}_k^l$.

Let us assume a model of m sources, n sensors, and a mixing matrix having constant scalar elements. The case of an instantaneous mixing matrix per (2.2) results in an Instantaneous Multiple-Input Multiple-Output (I-MIMO) model. If the mixing system is comprised of finite impulse response (FIR) filters instead of fixed constants, as described in (2.4), the result is a Finite Impulse Response Multiple-Input Multiple-Output (FIR-MIMO) model. This model is shown in Figure (2.1).

Several assumptions [30, 31] are needed for successful blind separation:

A1: The number of sensors n is greater than or equal to the number of sources m. This is a necessary assumption in most existing algorithms. However, it has been shown that in some applications, i.e. communications with finite symbol alphabet, the number of sources can be greater than the number of sensors [143].

A2: The source signals are mutually independent at each time instant k.

A3: At most one source is normally distributed. This a valid assumption only for the noise free model (2.1).

A4: The mixing matrix $\mathbf{A}$ *is full rank.* If the mixing matrix is comprised of FIR filters, which is the case of FIR-MIMO systems, then it admits an FIR left inverse, i.e. it is minimum phase.

A5: Sources have finite second moments.

In recent years, new methods and new underlying assumptions have been introduced. Among the previously mentioned assumptions, some of the hypotheses are application dependent.

A6: Sources are zero mean and stationary.

A7: A part of the sources are known at the receiver. This hypothesis is used in communications in the form of a *training sequence*.

A8: Sources have constant modulus. This property arises in the case of M-ary phase shift keying (M-PSK) sources.

A9: Sources have finite alphabet. This means that the source signals are chosen from a finite set such as $\pm$ BPSK or a set of phase shifts for DQPSK signal.

A10: The noise $\mathbf{v}$ *is white and Gaussian*.

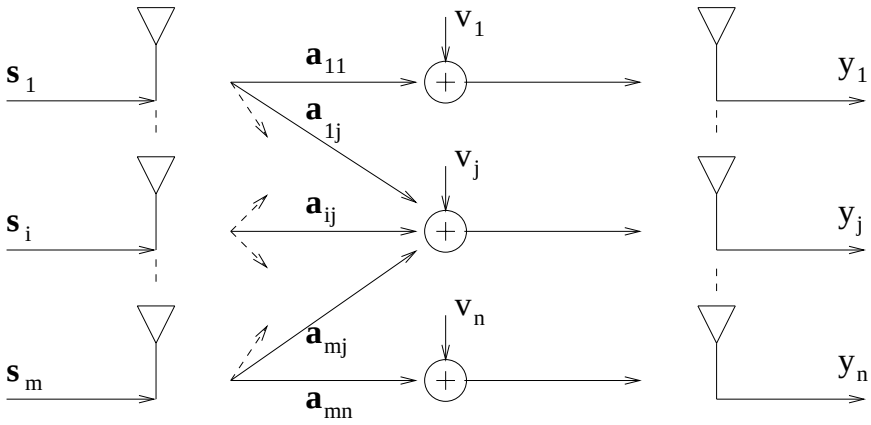

Figure 2.1: FIR-MIMO model.

The separation task at hand is to estimate the original source signals with high fidelity given noisy mixture measurements. In the case of I-MIMO model this is done by estimating either a separating matrix $\mathbf{W}$ or a mixing matrix $\hat{\mathbf{A}}$. An estimate $\mathbf{x}$ of unknown sources $\mathbf{s}$ is then given by

$$\hat{\mathbf{s}}_k = \mathbf{x}_k = \mathbf{W}\mathbf{y}_k = \hat{\mathbf{A}}^{-1}\mathbf{y}_k. \tag{2.5}$$

Some algorithms include a whitening stage prior to separation. During this stage the observed mixtures are spatially decorrelated and signal powers are normalized to unity. In addition, by projecting the input data along m signal subspace eigenvectors, the problem becomes easier to solve because the separating matrix will be an orthogonal matrix. Reducing the dimension of data from n to m is very important in some applications were n is much greater than m. For example, in the case of medical EEG and MEG measurements, this is often necessary because of the high dimensionality of the data. If the observations have been whitened, no inversion is needed to separate the sources since the separating matrix is orthogonal with $\hat{\mathbf{A}}^{-1} = \hat{\mathbf{A}}^{T}$. The sources may be recovered to within a permutation and a scaling, by matrices $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{S}}$ respectively:

$$\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{S}} = \mathbf{WA}. \tag{2.6}$$

The product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{S}}$ may be seen as a performance measure. Then a perfect separation leads to an identity matrix, $\mathbf{WA} = \mathbf{I}$.

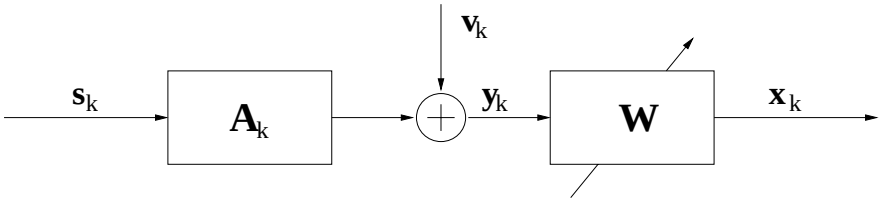

Figure 2.2: Adaptive I-MIMO model.

Source separation is a filtering problem that includes separation, deconvolution, or equalization [31]. Estimation of the mixing matrix $\mathbf{A}_k^l$, in the case of FIR-MIMO, is also known as channel identification. When *a priori* information such as training sequences is available, the process of source separation or channel identification is said to be *informed*. When no information about the sources or channel is known, such as a known training sequence of sufficient

length, the process of recovering the transmitted information or to identify the channel is said to be *blind*. If only limited knowledge about the sources is present, for example if the signal values are known part of the time, the processing is called *semi-blind* [31].

2.2 Key concepts in BSS

The main assumption of ICA is that the source signals $\mathbf{s}$ are independent. The initial m sources together generate an m-dimensional probability density function (*pdf*) $p(\mathbf{s})$. Statistical independence among the sources means that the joint source density factorises as:

$$p(\mathbf{s}) = \prod_{i=1}^{m} p(s_i(t)). \tag{2.7}$$

The same statement can be done for the separated sources $\mathbf{x}$. If the *pdf* of the estimated sources also factorises then they are independent.

The Kullback-Leibler divergence is a measure of the distortion between two probability density functions $p_{\mathbf{x}}(\mathbf{x})$ and $q_{\mathbf{x}}(\mathbf{x})$. The Kullback-Leibler divergence between $p_{\mathbf{x}}(\mathbf{x})$ and $q_{\mathbf{x}}(\mathbf{x})$ is given by:

$$\mathcal{K}(p,q) = \int p_{\mathbf{x}}(\mathbf{x}) \log \left(\frac{p_{\mathbf{x}}(\mathbf{x})}{q_{\mathbf{x}}(\mathbf{x})} \right) d\mathbf{x}. \tag{2.8}$$

This expression can be interpreted as a distance measure because it is always non-negative and is equal to zero only when $p_{\mathbf{x}}(\mathbf{x}) = q_{\mathbf{x}}(\mathbf{x})$. Due to this property, Kullback-Leibler divergence can be used to measure the mutual independence of output signals $\mathbf{x}$.

Estimation of a source model in blind separation usually involves formulating and then minimizing a contrast function [30]. Due to the fact that in practice finite data sets are available, the concept of the estimating function was introduced. These concepts will be briefly described in the next sections.

2.2.1 Contrast functions

Source separation can be obtained by optimizing a contrast function. These are real-valued functions of the distribution of the output $\mathbf{x}_k = \mathbf{W}\mathbf{y}_k$ and which must be designed such that separation is achieved when they reach their minimum value. In other words, using a contrast function turns

the source separation problem into an optimization problem. Typical optimization algorithms include gradient methods, Newton-type methods, and other techniques [65].

Contrast functions are based on entropy, mutual independence, higher-order decorrelations, or divergence between the joint distribution of $\mathbf{x}$ and a model. Important properties of the algorithms used to optimize the contrast functions include convergence speed, numerical stability, and memory requirements.

Likelihood

Let $\mathbf{r}$ denote a random vector with distribution q. The maximum likelihood principle is associated with a contrast function:

$$\phi_{ML}(p_{\mathbf{x}}) = \mathcal{K}(p_{\mathbf{x}}, p_{\mathbf{r}}). \tag{2.9}$$

This means that we have to find a mixing matrix $\mathbf{A}$ such that the distribution of the separated sources $\mathbf{x}$ is as close as possible (in the Kullback divergence sense) to the hypothesized distribution of the original sources. One problem of ML contrasts is that if the hypothesized distributions of the sources are not correct we will not obtain the desired results. Obviously, in blind separation the source distributions are unknown.

Mutual Information

In the case of mutual information, the idea is to minimize $\mathcal{K}(p_{\mathbf{x}}, p_{\mathbf{r}})$ with respect to $\mathbf{A}$ taking into account the distribution of $\mathbf{x}$ and with respect to the model distribution $\mathbf{r}$. Let $\check{\mathbf{x}}$ denote a random vector with independent entries and with each entry distributed in the same way as the corresponding entry of $\mathbf{x}$. We obtain:

$$\mathcal{K}(p_{\mathbf{x}}, p_{\mathbf{r}}) = \mathcal{K}(p_{\mathbf{x}}, p_{\check{\mathbf{x}}}) + \mathcal{K}(p_{\check{\mathbf{x}}}, p_{\mathbf{r}}). \tag{2.10}$$

The last term is minimized by taking $\mathbf{r} = \check{\mathbf{x}}$ for which $\mathcal{K}(p_{\check{\mathbf{x}}}, p_{\mathbf{r}}) = 0$. The contrast function is:

$$\phi_{MI}(p_{\mathbf{x}}) = \mathcal{K}(p_{\mathbf{x}}, p_{\check{\mathbf{x}}}) \tag{2.11}$$

which can be interpreted as the Kullback divergence between a distribution and the closest distribution with independent entries.

Orthogonal contrasts

These contrasts are used when the data has been prewhitened. In such cases, the minimization of the contrast function must take place under the constraint that $\mathrm{E}\left[\mathbf{x}\mathbf{x}^T\right] = \mathbf{I}$, where $\mathrm{E}\left[\cdot\right]$ is the expectation operation. The mutual information contrast function becomes:

$$\phi^o_{MI}(p_{\mathbf{x}}) = \sum_i H\left[\mathbf{x}_i\right], \tag{2.12}$$

where $H\left[\cdot\right]$ is the entropy. In other words, minimizing the mutual information between the entries of $\mathbf{x}$ is equivalent to minimizing the sum of the entropies of the entries of $\mathbf{x}$.

Cumulants

Higher Order Statistics (HOS) can be used to define contrast functions. Higher order information may be expressed by cumulants. Given the zero-mean vector $\mathbf{x}$, the most relevant cumulants for BSS are those of second and fourth order [21], defined as:

$$\mathrm{C}_{ij}\left[p_{\mathbf{x}}\right] \stackrel{\text{def}}{=} \mathrm{E}\left[x_i, x_j\right] \tag{2.13}$$

and as:

$$\begin{aligned}\mathrm{C}_{ijkl}\left[p_{\mathbf{x}}\right] \stackrel{\text{def}}{=}\ & \mathrm{E}\left[x_i, x_j, x_k, x_l\right] - \mathrm{E}\left[x_i, x_j\right]\mathrm{E}\left[x_k, x_l\right] - \\ & \mathrm{E}\left[x_i, x_k\right]\mathrm{E}\left[x_j, x_l\right] - \mathrm{E}\left[x_i, x_l\right]\mathrm{E}\left[x_j, x_k\right].\end{aligned} \tag{2.14}$$

The third order cumulants are:

$$\mathrm{C}_{ijk}\left[p_{\mathbf{x}}\right] \stackrel{\text{def}}{=} \mathrm{E}\left[x_i, x_j, x_k\right]. \tag{2.15}$$

Under the assumption of independence, the cross entries of the sources are zero, and:

$$\mathrm{C}_{ii}\left[p_{\mathbf{s}}\right] = \mathrm{E}\left[s_i^2\right] \stackrel{\text{def}}{=} \sigma_i^2, \tag{2.16}$$

where σ_i^2 is the variance of the ith source. Similarly,

$$\mathrm{C}_{iiii}\left[p_{\mathbf{s}}\right] = \mathrm{E}\left[s_i^4\right] - 3\mathrm{E}\left[s_i^2\right]^2 \stackrel{\text{def}}{=} \kappa_i, \tag{2.17}$$

where κ_i is the kurtosis of the ith source. Signals with positive kurtosis, (the tails of their densities decay more slowly than the Gaussian density and are sharply peaked around their mean) are known as super-Gaussian. Signals with negative kurtosis (rapidly decaying tails) are called sub-Gaussian.

The likelihood contrast ϕ_{ML} is a measure of mismatch between an output distribution and a model source distribution. A cruder measure can be defined from the quadratic mismatch between the cumulants:

$$\phi_2\left(p_{\mathbf{x}}\right) = \sum_{ij}\left(\mathrm{C}_{ij}\left(p_{\mathbf{x}}\right) - \mathrm{C}_{ij}\left(p_{\mathbf{s}}\right)\right)^2 = \sum_{ij}\left(\mathrm{C}_{ij}\left(p_{\mathbf{x}}\right) - \sigma_i^2\delta_{ij}\right)^2 \tag{2.18}$$

and

$$\phi_4\left(p_{\mathbf{x}}\right) = \sum_{ijkl}\left(\mathrm{C}_{ijkl}\left(p_{\mathbf{x}}\right) - \mathrm{C}_{ijkl}\left(p_{\mathbf{s}}\right)\right)^2 = \sum_{ijkl}\left(\mathrm{C}_{ijkl}\left(p_{\mathbf{x}}\right) - \kappa_i\delta_{ijkl}\right)^2, \tag{2.19}$$

where δ is the Kronecker symbol. Cardoso [21] pointed out that the measure defined by (2.18) is not a true contrast in the BSS sense, as it reaches zero when $\mathbf{x}$ is linearly decorrelated. The use of the fourth order information ϕ_4 leads to independence.

If $\mathbf{s}$ and $\mathbf{x}$ are symmetrically distributed with distributions that are close to normal, then the maximum likelihood approach can be approximated as [21]:

$$\begin{aligned}\phi_{ML} &= \mathcal{K}\left(p_{\mathbf{x}}, p_{\mathbf{s}}\right) \approx \phi_{24}(p_{\mathbf{x}}) \\ &\stackrel{\text{def}}{=} \frac{1}{48}\left(12\phi_2(p_{\mathbf{x}}) + \phi_4\left(p_{\mathbf{y}}\right)\right).\end{aligned} \tag{2.20}$$

If the kurtosis values of all sources have the same sign, the sum of the fourth moments can be used as a contrast function [94]:

$$\phi^o_{\text{moreau}}(p_{\mathbf{x}}) \stackrel{\text{def}}{=} \sum_{i=1}^{m}\mathrm{E}\left[x_i^4\right]. \tag{2.21}$$

Cardoso proposed to test the independence on a smaller subset of cross-cumulants [24]. His approach resulted in the Joint Approximate Diagonalization of Eigen-matrices (JADE) technique, which, under the whiteness constraint, has the contrast function:

$$\phi^o_{JADE} \stackrel{\text{def}}{=} \sum_{ijkl\neq ijkl}\mathrm{C}^2_{ijkl}\left(p_{\mathbf{x}}\right). \tag{2.22}$$

2.2.2 Score functions

Choosing the *score* or *squashing* functions is very important since they describe the source model. The score functions $\varphi_1, \ldots, \varphi_m$ are defined as the log derivatives of the source densities $q_1, \ldots, q_m$:

$$\varphi_i \triangleq -\left(\log q_i\right)' \quad \text{or} \quad \varphi\left(\cdot\right) = -\frac{q_i\left(\cdot\right)'}{q_i\left(\cdot\right)}. \tag{2.23}$$

In the case of a zero-mean unit-variance Gaussian variable s with density $q(s) = (2\pi)^{-1/2} \exp -(s^2/2)$, the associated score function is $\varphi\left(s\right) = s$. Gaussian densities are associated with linear score functions. Non-Gaussian modeling results in considering non-linear score functions [21].

Several approximations for the score functions have been used in the literature. For example, Bell and Sejnowski [17] used a fixed source model assuming that all the initial sources have the same kurtosis. This type of processing was further developed by Girolami [49], who used different score functions for sub- and super-Gaussian sources in order to separate mixtures of densities. Based on the stability analysis introduced in [8], Douglas [39] proposed switching between nonlinearities by analyzing the statistics on each output channel. Other approaches do exist for selecting the score functions [78]. Generalized exponentials or mixtures of Gaussians have also been used to model sources. See [112] for a review and [90] for a more detailed analysis.

2.2.3 Estimating functions

Due to their design, all contrast functions reach their minimum at a separating point. However, in practice contrast functions are estimated from a finite data set. Thus, the sample-based contrasts depend on the sample distribution of $\mathbf{x}$. Due to the errors introduced by estimation using a small data set, statistical characterization of the minima of sample-based contrasts is needed. In this sense the notion of an estimation function was introduced [21]. The estimation function for blind separation is a function $\mathcal{F} : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$. Considering a batch of N samples, the estimation function $\mathcal{F}$ is associated with an estimating equation:

$$\frac{1}{N} \sum_{l=1}^{N} \mathcal{F}\left[\mathbf{x}(l)\right] = 0. \tag{2.24}$$

The gradient of the ML contrast function was found by Pham [107]:

$$\nabla \phi_{ML}\left(p_{\mathbf{x}}\right) = \mathrm{E}\left[\mathcal{F}_{\varphi}\left(\mathbf{x}\right)\right], \tag{2.25}$$

where

$$\mathcal{F}_{\varphi}\left(\mathbf{x}\right) \stackrel{\text{def}}{=} \varphi\left(\mathbf{x}\right)\mathbf{x}^{T} - \mathbf{I} \tag{2.26}$$

and $\mathbf{I}$ is an identity matrix. The ML contrast function achieves its minimum at points where its relative gradient cancels, i.e. at the points which are solutions of the equation $\mathrm{E}\left[\mathcal{F}_{\varphi}\left(\mathbf{x}\right)\right] = 0$. We note that ML estimates correspond exactly to the solution of an estimating equation [21].

Under the whiteness constraint, the ML estimation function is:

$$\mathcal{F}^{o}_{\varphi}\left(\mathbf{x}\right) \stackrel{\text{def}}{=} \mathbf{x}\mathbf{x}^{T} - \mathbf{I} + \varphi\left(\mathbf{x}\right)\mathbf{x}^{T} - \mathbf{x}\varphi\left(\mathbf{x}\right)^{T}. \tag{2.27}$$

The estimating function for the orthogonal contrast $\phi^{o}_{\text{moreau}}\left(p_{\mathbf{x}}\right)$ given in (2.21) has the same form as given in (2.27) but with $\varphi_i\left(x_i\right) = x_i^3$. Not all the contrast functions have estimating functions which can be expressed in the form (2.24). However, one can often find an asymptotic estimating function in the sense that the solution of the associated estimating equation

is very close to the minimizer of the estimated contrast [21].

2.3 Different classes of algorithms

The ways in which separation algorithms process the data can be used as basic classification criteria. There are situations when the whole data set is available. In such case the processing is done in batch mode [24, 63, 64] considering the whole set of available samples. Algorithms of this type are *batch* algorithms. In real-time applications the data is available one at the time, meaning that at each time index k we receive a n-dimensional vector of observations $\mathbf{y}_k$. Based on the new received data vector and possibly on a vector of some previously-received data, the task is to estimate the initial sources. Algorithms of this type are *on-line* algorithms [23].

The advantage of on-line algorithms is that they enable faster adaptation in a time-varying environment due to the fact that the input $\mathbf{y}_k$ can be used in the algorithm immediately. A resulting trade-off is that the convergence may be slow and the convergence rate may depend on the choice of the learning rate. A bad choice of the learning rate can lead to very poor results.

Batch algorithms should be used in situations where fast real-time adaptation is not necessary [64] and in which mixing systems or source statistics are not time-varying.

Existing blind separation algorithms can be divided into two main groups. Methods in the first group attempt to find a separation matrix directly, while the methods from the second group use whitening before determining a separation matrix. Whitening has some advantages such as reduction of the data dimension from n to m and also noise attenuation. The separation task is made easier because the components of the whitened vectors $\mathbf{v}$ are already uncorrelated and we have to search for an orthogonal separating matrix. Moreover, using real-world data it has been shown [48] that whitening can improve both the convergence speed and the separation performance. A good example is the application of BSS to anti-personnel land mine detection [77]. In this case, blind separation is used to detect the anti-personnel land mines based on a set of sensor signal measurements. The number of mixtures is very high in comparison with the number of sources, i.e. there may be $n = 2651$ mixtures and $m = 22$ sources. Thus, it is impractical to apply algorithms which search for a separation (or mixing) matrix without prior whitening of the sources. Whitening has also some disadvantages. For example, if some of the source signals are very weak or the mixture matrix is ill-conditioned, prewhitening may greatly reduce the accuracy of the algorithm.

Different techniques for recovering the transmitted sources have been proposed. One of the first algorithms that appeared in the literature was proposed by Herault and Jutten [58]. The algorithm is based on the idea of measuring independence of the separated sources by pairwise nonlinear decorrelation. The mixing model of equation (2.1) is used. If the sources are zero mean and have symmetric densities, and if the selected nonlinearities f and g are odd, then the expectation $\mathrm{E}[f(s_i)g(s_j)]$ is zero.

The family of gradient-based algorithms is very important in the BSS literature. Bell and Sejnowski [17] derived an algorithm based on maximizing the entropy of a nonlinear output. The algorithm uses a stochastic gradient optimization method without prewhitening, and successfully separates speech sources. Amari et al. [10] proposed an improvement to Bell's stochastic gradient algorithm, based on using the natural gradient. The goal is to update a separation matrix in the direction of the natural gradient [7], which leads to faster convergence than with stochastic

gradient algorithms. A similar algorithm, called relative gradient, was independently proposed by Cardoso and Laheld [23]. This algorithm has also the equivariance property, meaning that its behavior does not depend on the nature of the mixing matrix. Amari proposed a similar algorithm based on minimizing the mutual information using natural gradient learning [10].

Another class of algorithms used to optimize contrast functions is represented by Jacobi algorithms. They are called Jacobi due to the fact that the goal is to maximize measures of independence by a technique akin to that of the Jacobi method of diagonalization. The Jacobi method is an iterative technique of optimization over the set of orthonormal matrices which are obtained as a sequence of plane rotations. Several algorithms have been proposed. The first one was introduced by Comon [30]. As pointed out in [22], this is a data-based algorithm, meaning that it works through a sequence of Jacobi sweeps on whitened data until a given contrast is optimized. A statistic-based algorithm, JADE, was introduced by Cardoso [24] where the plane rotations are applied to the cumulant matrices, instead of to the data itself. A mixed approach, called SHIBBS (SHIfted Blocks for Blind Separation), was introduced in [22] where the update to get the separated sources is made on the data itself and the rotation matrix that is applied to the data is computed in a statistic-based procedure. One advantage of Jacobi algorithms is that no tuning is needed (in their basic versions) as opposed to the gradient-based algorithms in which a learning schedule is necessary and usually implemented in a heuristic manner [22].

Stemming from Principal Component Analysis, the class of nonlinear PCA algorithms is also of great importance [73]. Several algorithms were introduced, for instance that of [72]. It has been shown that nonlinear PCA can separate signals in the presence of a noisy time-varying mixing model [74, 75]. The connections between several ICA algorithms, such as the Bell-Sejnowski [17] algorithm or the EASI algorithm[23], and information-theoretic contrasts have been shown by Karhunen et al. [76]. An overview of adaptive algorithms is given by Amari et al. [9], and a detailed description of statistical principles used in BSS is made by Cardoso [21].

2.4 Blind Deconvolution

Plenty of research has been done in the area of blind deconvolution [9, 12, 54, 57]. Various scenarios have been considered, starting from the single-input multiple-output (SIMO) model obtained by oversampling at the receiver or by using several receivers [62], to the MIMO case [12, 54]. Different techniques based on Higher Order Statistics [131], subspace decomposition [89] or multichannel frequency-domain deconvolution [84] have been reported in the literature. Typically assumptions include linear time-invariant (LTI) systems, infinite SNR, and infinite equalizer length [131]. In contrast, real systems are time-varying (TV), SNR values are low, and equalizer lengths are finite.

Using the model described by equation (2.4), the goal of blind deconvolution is to estimate source signals using a multichannel linear filter of the form:

$$\mathbf{x}_k = \sum_{l=0}^{L_r} \mathbf{W}_k^l \mathbf{y}_{k-l}, \tag{2.28}$$

where $\mathbf{W}_k^l$ are the $(m \times n)$ matrix coefficients of the separation system and L_r is a filter length parameter.

A very interesting technique for possibly transforming the BSS algorithms into multichannel blind deconvolution algorithms is presented in [38, 40]. This is based on assuming that $m = n$ and that the mixing matrix $\mathbf{A}$ is circulant. A circulant matrix is completely specified by any one row or column, as the other rows or columns of the matrix are simply modulo-shifted versions of this row [38]. For example, the first column of $\mathbf{A}$ is $[a_0 \;\; a_1 \;\; \ldots \;\; a_{m-1}]^T$, the second column of $\mathbf{A}$ is $[a_{m-1} \;\; a_0 \;\; \ldots \;\; a_{m-2}]^T$, and so on. However, a principal problem is that the assumption of a circulant matrix is artificial. No physical mixing system exhibits this structure [38].

The presentation of blind deconvolution techniques based on this transformation is beyond the scope of this discussion. We will simply state that the transformation process involves following three rules that make associations between matrices in the BSS task, such as $\mathbf{A}$, $\mathbf{W}_k$ and matrix sequences in the multichannel blind deconvolution task such as $\mathbf{A}^l$, $\mathbf{W}_k^l$. These three rules can be summarized as follows [38, 40]:

- Multiplication of two matrices in instantaneous BSS (I-BSS) is equivalent to convolution of their associated matrix sequences in multichannel blind deconvolution (MBD).

- Addition of two matrices in I-BSS is equivalent to element by element addition of their associated matrix sequences in MBD.

- Transposition of a matrix in I-BSS is equivalent to element by element transposition and time reversal of its associated matrix sequence in MBD.

The previous procedure can be applied to the density matching BSS algorithm using natural gradient adaptation [12, 38] and to contrast function optimization for I-BSS algorithms.

2.5 Applications of BSS

Due to the volume of research on blind separation over the past years, the recognized applications of BSS are numerous. For example, in *brain imaging* applications we may capture recordings of electric (electroencephalograms, EEG) and magnetic (magnetoencephalograms, MEG) fields of signals emerging from neural currents within the brain. It is important to extract the essential features from the data allowing a better representation and understanding of their properties. An important application of BSS is the separation of artifacts from EEG and MEG data [133].

In *wireless communications* an essential issue is the sharing of the common transmission medium among several users. In Code Division Multiple Access (CDMA) systems all users occupy the same frequency band simultaneously. The users are identified via unique codes. During the transmission different users' signals become mixed, the user can be identified from the mixture by applying the code at the receiver. In downlink (mobile phone) signal processing each user knows only one code. The codes of the other users are unknown. By modeling the CDMA signal as a linear combination of convolved independent symbol sequences [33], BSS techniques can be applied for the separation of the sources. I-MIMO model is used for narrowband communications applications and FIR-MIMO for frequency selective channels. Another communications application can be in GSM where blind separation can be used to achieve blind equalization under some conditions [144].

Speech separation is an important and attractive application domain for BSS. One of the main applications is the separation of simultaneous audio sources in reverberating or echoing environments, i.e. inside a room. Speech enhancement is a very desirable application where only one

signal is of interest and the rest are considered to be nuisance signals [50]. Enhancement of voice quality in mobile phones, would be one important application, especially in car environments. In real environments multiple paths are common and blind deconvolution is usually necessary. Torkkola [127] gives an extensive survey, with many references, of blind separation of audio signals.

Many other applications exist [65]. In fact, every application which leads to the BSS model can be of interest even if the assumptions about the BSS model are not so close. In many practical BSS applications, observations are noisy, the mixing system and/or source statistics may be time-varying, and source signals may appear and disappear randomly. Moreover, the delays associated with batch processing may be intolerable. Hence, it is important to develop recursive separation algorithms that take into account both noise and the time-varying nature of the problem and that allow real-time computation.

Chapter 3

Adaptive Whitening

3.1 Introduction

Several BSS algorithms use *whitening* transform prior to performing separation. Due to this operation, the search for a separating matrix is reduced to the search for an orthogonal separating matrix. The number of unknown parameters which must be determined is reduced as well. By definition, $\mathbf{V}_k$ is a whitening matrix if the outputs $\mathbf{z}_k = \mathbf{V}_k\mathbf{y}_k$ are spatially white, i.e. $\mathrm{E}\left[\mathbf{z}_k\mathbf{z}_k^T\right] = \mathbf{R}_\mathbf{z} = \mathbf{I}$. Adaptive whitening consists of updating the matrix $\mathbf{V}_k$ such that it converges to a point where $\mathbf{R}_\mathbf{z} = \mathbf{I}$. The covariance matrix $\mathbf{R}_\mathbf{y}$ may be time-varying if the mixing matrix and/or the sources are time-varying. The ability to adapt is needed if $\mathbf{R}_\mathbf{y}$ varies, otherwise, a recursive formula has only computational advantages. Let us consider the noisy model from (2.2). When the number of sensors n is greater than the number of sources m, the covariance matrix $\mathbf{R}_\mathbf{y}$ is given by:

$$\mathbf{R}_\mathbf{y} = \mathbf{A}\mathbf{R}_\mathbf{s}\mathbf{A}^T + \sigma_v^2\mathbf{I}, \tag{3.1}$$

where $\mathbf{R}_\mathbf{s} = \mathrm{E}\left[\mathbf{s}\mathbf{s}^T\right]$ and σ_v^2 is the noise variance. Using matrices of eigenvectors and eigenvalues, $\mathbf{R}_\mathbf{y}$ can be written as:

$$\mathbf{R}_\mathbf{y} = \mathbf{E}_\mathbf{s}\mathbf{D}_\mathbf{s}\mathbf{E}_\mathbf{s}^T + \mathbf{E}_\mathbf{v}\mathbf{D}_\mathbf{v}\mathbf{E}_\mathbf{v}^T, \tag{3.2}$$

where $\mathbf{D}_\mathbf{s}$ is an $m \times m$ diagonal matrix given by $\mathbf{D}_\mathbf{s} = \mathrm{diag}\left[\lambda_1, \ldots, \lambda_m\right]$, with λ_m denoting the mth largest eigenvalues of the covariance matrix $\mathbf{R}_\mathbf{y}$, $\mathbf{E}_\mathbf{s}$ is an $n \times m$ matrix containing the $\mathbf{e}_i$ principal eigenvectors, $\mathbf{E}_\mathbf{s} = \left[\mathbf{e}_1, \ldots, \mathbf{e}_m\right]$ corresponding to the m largest eigenvalues.

The subspace spanned by the eigenvectors contained in $\mathbf{E_s}$ is the signal subspace. The matrix $\mathbf{D}_v = \text{diag}\,[\lambda_{m+1}, \ldots, \lambda_n]$ is diagonal and contains the remaining $n - m$ noise eigenvalues and $\mathbf{E_v} = [\mathbf{e}_{m+1}, \ldots, \mathbf{e}_n]$ contains the corresponding $n-m$ noise eigenvectors. The subspace spanned by the noise eigenvectors is the noise subspace. The signal subspace is orthogonal to the noise subspace.

The dimension of the signal subspace can be estimated by inspecting the eigenvalues of $\mathbf{R_y}$. The signal subspace eigenvalues usually exhibit a clear pattern. They are a linear combination of the source powers $\mathrm{E}\,[s_k(i)^2]$ added the noise power σ_v^2 [73]. If the signal to noise ratio (SNR) is high enough, the m largest eigenvalues are much larger than the other $n - m$ eigenvalues. At low SNR this pattern may not be so clear, and information theoretical criteria such as minimum description length (MDL) can be used to find the number of signals [134].

The chapter presents adaptive methods for performing whitening. The techniques addressed are based on Principal Component Analysis. Serial update of whitening matrices is also presented. We also discuss how changes in the dimension of the signal subspace can be tracked in real time.

3.2 Subspace Tracking

Since whitening is essentially a decorrelation followed by scaling, Principal Component Analysis (PCA) can be used [65]. Considering the covariance matrix $\mathbf{R_y} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, where $\mathbf{E}$ is a $n \times n$ matrix of eigenvectors and $\mathbf{D}$ is a $n \times n$ diagonal matrix of eigenvalues, the whitening matrix $\mathbf{V}$ is given by:

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T. \tag{3.3}$$

If more sensors than sources are present in the system, the whitening matrix is formed using the signal subspace. Thus $\mathbf{V_s} = \mathbf{D_s}^{-1/2}\mathbf{E_s}^T$. This reduces the dimension of the data from n to m and allows the search for an orthogonal separating matrix of dimension $m \times m$. The whitening matrix $\mathbf{V}$ introduced in (3.3) is not a unique whitening matrix. Let consider $\boldsymbol{\mathcal{U}}$ an orthogonal matrix and

let apply the transform $\mathcal{U}\mathbf{V}$ to the received data $\mathbf{y}_k$. Then, we obtain:

$$\begin{aligned} \mathrm{E}\left[\mathbf{z}_k\mathbf{z}_k^T\right] &= \mathcal{U}\mathbf{V}\mathrm{E}\left[\mathbf{y}_k\mathbf{y}_k^T\right]\mathbf{V}^T\mathcal{U}^T \\ &= \mathcal{U}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{E}\mathbf{D}\mathbf{E}^T\mathbf{E}\mathbf{D}^{-1/2}\mathcal{U}^T \\ &= \mathcal{U}\mathbf{I}\mathcal{U}^T = \mathbf{I}, \end{aligned} \tag{3.4}$$

where we have used the orthogonality of the eigenvectors, i.e. $\mathbf{e}_i^T\mathbf{e}_j = 0$ for $i \neq j$. Thus, any matrix $\mathcal{U}\mathbf{V}$, with $\mathcal{U}$ an orthogonal matrix, is also a whitening matrix.

If the covariance matrix of the received data varies in time, on-line update of the eigenvalues λ_i and eigenvectors $\mathbf{e}_i$ is needed in order to accurately track/adjust the whitening matrix $\mathbf{V}_k$ at each time step k. A good review of methods for tracking principal singular values and vectors is given by Comon and Golub [32].

Any subspace tracking algorithm which exhibits good convergence and tracking capability can be used for adaptive whitening. Some subspace tracking algorithms track only the eigenvectors. Using these results in decorrelated data with a possibly-incorrect scale. One BSS algorithm that performs only decorrelation and then employs a contrast based approach to find a separation matrix was introduced by Douglas [36]. However, some of the BSS algorithms require that the data has unit power. Thus an estimate of the eigenvalues is also needed in order to update the whitening matrix $\mathbf{V}_k$.

3.2.1 PAST and PASTd

Yang [136] proposed an on-line algorithm for tracking the m-dimensional principal signal subspace by using an approximate Recursive Least Squares (RLS) type of update. The Projection Aproximation Subspace Tracking (PAST) algorithm computes a subspace eigenvector matrix estimate $\mathbf{E}_s$ that minimizes the least squares criterion:

$$\mathcal{J}_{PAST}(k) = \sum_{l=1}^{k} \gamma^{k-l} \parallel \mathbf{y}_l - \mathbf{E}_\mathbf{s}\mathbf{z}_l \parallel^2, \tag{3.5}$$

where $\parallel \cdot \parallel$ denotes the Euclidean norm and γ is a forgetting factor needed in tracking time-varying system. For notational convenience we will use $\mathbf{E}$ instead of $\mathbf{E}_\mathbf{s}$ in the following. It is also assumed that we have more sensors than sources, i.e. $n > m$. The recursive update for the

$n \times m$ matrix $\mathbf{E}$ at time k is:

$$\begin{aligned}
\mathbf{z}_k &= \mathbf{E}_{k-1}^T \mathbf{y}_k \\
\mathbf{p}_k &= \mathbf{y}_k - \mathbf{E}_{k-1}\mathbf{z}_k \\
\mathbf{k}_k &= \frac{\mathbf{P}_{k-1}\mathbf{z}_k}{\gamma + \mathbf{z}_k^T \mathbf{P}_{k-1}\mathbf{z}_k} \\
\mathbf{P}_k &= \frac{1}{\gamma}\mathrm{Tri}\left\{\mathbf{P}_{k-1} - \mathbf{k}_k \mathbf{z}_k^T \mathbf{P}_{k-1}^T\right\} \\
\mathbf{E}_k &= \mathbf{E}_{k-1} + \mathbf{p}_k \mathbf{k}_k^T,
\end{aligned} \tag{3.6}$$

where $\mathbf{p}_k$ is the $n \times 1$ *a priori* estimation error vector (or the innovation), $\mathbf{k}_k$ is the $m \times 1$ gain vector and $\mathbf{P}_k$ is the $m \times m$ inverse of the correlation matrix $\mathbf{R}_{\mathbf{z}}$. The notation Tri means that only the upper triangular part of the argument is computed. Its transpose is copied to the lower triangular part so that the resulting matrix becomes symmetric.

The forgetting factor $0 < \gamma \leq 1$ allows tracking when the system operates in a non-stationary environment. The value $\gamma = 1$ corresponds to the standard least square solution. The choice of initial values for $\mathbf{E}$ and $\mathbf{P}$ affects the transient behavior but not the steady state performance of the algorithm [136]. In order to avoid transient behavior problems $\mathbf{P}(0)$ must be a Hermitian positive definite matrix and $\mathbf{E}(0)$ should contain m orthonormal vectors. These matrices can be calculated, for example, from an initial block of data.

The PAST algorithm does not maintain the orthonormality of the estimate $\mathbf{E}_k$ during the adaptation [136]. Douglas proposed [37] a modification of PAST that enforces the constraint $\mathbf{E}_k \mathbf{E}_k^T = \mathbf{I}$. This algorithm employs m identical Householder transformations to update $\mathbf{E}_k$. For adaptive subspace analysis, the general Householder-based update is:

$$\boldsymbol{\mathcal{E}}_k = \boldsymbol{\mathcal{E}}_{k-1}\left[\mathbf{I} - \frac{\mathbf{m}_k \mathbf{m}_k^T}{\| \mathbf{m}_k \|}\right], \tag{3.7}$$

where $\boldsymbol{\mathcal{E}}_{k-1}$ is a matrix whos columns have to be rotated and $\mathbf{m}_k$ is the Householder vector. The Householder-based update for $\mathbf{E}_k$ behaves similarly to PAST within the constraint space $\mathbf{E}_k \mathbf{E}_k^T = \mathbf{I}$. The modifications are as follows:

$$\begin{aligned}
\mathbf{m}_k &= \mathbf{p}_k - \frac{1}{2} \| \mathbf{p}_k \|^2 \mathbf{E}_{k-1}\mathbf{k}_k \\
\mathbf{E}_k &= \mathbf{E}_{k-1} + \frac{\mathbf{m}_k \mathbf{k}_k^T}{1 + \frac{1}{4} \| \mathbf{p}_k \|^2 \| \mathbf{k}_k \|^2},
\end{aligned}$$

where $\mathbf{z}_k$, $\mathbf{k}_k$ and $\mathbf{P}_k$ are computed as given in (3.6).

Based on the deflation technique, Yang [136] introduced PASTd algorithm derived from PAST. PASTd enables also the sequential estimation of the eigencomponents. The key idea in this technique is the following. First the most dominant eigenvector is updated by applying PAST algorithm with $m = 1$. Then, the projection of the current data vector $\mathbf{y}_k$ along this eigenvector is removed from the current data vector. At this stage, the second dominant eigenvector becomes the most dominant eigenvector and can be extracted in the same manner. Repeating this procedure, all desired eigencomponents are estimated sequentially. The algorithm may be summarized as follows:

$$\begin{aligned}
\mathbf{y}_k^1 &= \mathbf{y}_k \\
\text{for} \quad i &= 1, 2, \ldots, m \\
z_k^i &= \left(\mathbf{e}_{k-1}^i\right)^T \mathbf{y}_k^i \\
b_k^i &= \gamma b_{k-1}^i + |z_k^i|^2 \\
\mathbf{p}_k^i &= \mathbf{y}_k^i - \mathbf{e}_{k-1}^i z_k^i \\
\mathbf{e}_k^i &= \mathbf{e}_{k-1}^i + \mathbf{p}_k^i \left[\frac{(z_k^i)^*}{b_k^i}\right] \\
\mathbf{y}_k^{i+1} &= \mathbf{y}_k^i - \mathbf{e}_k^i z_k^i,
\end{aligned} \tag{3.8}$$

where $\mathbf{e}_k^i$ is an estimate of the ith eigenvector of $\mathbf{R}_\mathbf{y}$ and b_k^i is an exponentially weighted estimate of the corresponding eigenvalue. The drawbacks of PASTd are the fact that it loses the orthonormality between $\mathbf{e}_i$ and a slightly increased complexity if $n \gg m$.

3.2.2 Subspace tracking by subspace averaging

Karasalo [70] proposed an algorithm for updating the covariance matrix by signal subspace averaging. A similar method was also proposed by Tufts et. al. [130]. The algorithm operates as follows [70]. The received data vector $\mathbf{y}_k$ is split into signal and noise subspaces:

$$\mathbf{y}_k = \mathbf{E}_{k-1}\mathbf{r}_k + \mathbf{q}_k c_k, \tag{3.9}$$

where $\mathbf{E}_{k-1}\mathbf{r}_k$ define the old signal subspace, $\mathbf{q}_k$ is part in orthogonal subspace and c is a normalization scalar. This decomposition involves computing the following variables:

$$\mathbf{r}_k = \mathbf{E}_{k-1}^H \mathbf{y}_k \tag{3.10}$$

$$\mathbf{q}_k = \mathbf{y}_k - \mathbf{E}_{k-1}\mathbf{r}_k \tag{3.11}$$

$$c_k = \parallel \mathbf{q}_k \parallel, \mathbf{q}_k = \mathbf{q}_k / c_k. \tag{3.12}$$

A major computational advantage stems from constructing a smaller $(m+1)\times(m+2)$ matrix $\mathbf{B}$ which preserves the properties of the $n \times n$ covariance matrix $\mathbf{R_y}$ and contains all the information needed to compute both the squared singular values λ_i and the associated eigenvectors $\mathbf{e}_i$.

$$\mathbf{B} = \begin{bmatrix} \sqrt{\beta_k}\mathbf{D}_{k-1} & 0 & \sqrt{\alpha_k}\mathbf{r}_k \\ 0 & \sqrt{\beta_k}\sigma_{k-1} & \sqrt{\alpha_k}c_k \end{bmatrix}, \tag{3.13}$$

where $\mathbf{D}_{k-1}$ is a diagonal matrix containing the square roots of the m principal eigenvalues estimated at time step $k-1$, σ_{k-1} is the square root of the noise variance, α_k and β_k are weighting coefficients. A singular value decomposition (SVD) of the matrix $\mathbf{B}$ must be performed:

$$\mathbf{\Gamma\Sigma\Lambda}^H = \mathbf{B}. \tag{3.14}$$

Finally, eigenvector and eigenvalue estimates are updated. The square roots of the eigenvalues, λ_k, are found in the upper left corner in $\mathbf{\Sigma}$. The corresponding eigenvectors are the m first columns in:

$$\begin{bmatrix}\mathbf{E}_{k-1} & \mathbf{q}_k\end{bmatrix}\mathbf{\Gamma}. \tag{3.15}$$

An update of the noise variance σ^2 is also obtained:

$$\sigma^2{}_k = \frac{1}{n-m}\left[\sigma^2_{m+1} + (n-m-1)\,\beta_k \sigma^2{}_{k-1}\right], \tag{3.16}$$

where σ^2_{m+1} is the $m+1$ diagonal element of $\mathbf{\Sigma}$. The weight coefficients α_k and β_k are very important in the process of tracking the principal subspace. The better the estimates of these coefficients, the closer the covariance matrix $\hat{\mathbf{R}}_\mathbf{y}$ is to the local true covariance matrix $\mathbf{R_y}$. In [70], the update of the weight coefficients α_k and β_k is done in such a way that the algorithm initially relies on the observed data and later relies more on the computed eigen-decomposition than on new received data.

One may be also interested in an update of the complete covariance matrix. It is obtained as follows:

$$\hat{\mathbf{R}}_{\mathbf{y}} = \hat{\mathbf{E}}_k \hat{\mathbf{D}}_k \hat{\mathbf{E}}_k^H + \hat{\sigma}_k^2 \mathbf{I}. \tag{3.17}$$

where the estimates $\hat{\mathbf{E}}_k$, $\hat{\mathbf{D}}_k$ and $\hat{\sigma}_k^2$ are found using the previous algorithm. The computational complexity of the previous algorithm is relatively low. However, one SVD is involved in the update process. The dimension depends on the dimension of the signal subspace. Hence, substantial savings are obtained if the dimension of signal subspace is small compared to the dimension of the data covariance matrix.

3.2.3 Serial update of the whitening matrix

Another method for adaptive whitening was proposed by Cardoso. In [23] the serial update of the whitening matrix is based on minimizing the 'distance' between $\mathbf{R}_{\mathbf{z}}$ and $\mathbf{I}$. The Kullback-Leibler divergence between two normal distributions with covariance matrices $\mathbf{R}_{\mathbf{z}}$ and $\mathbf{I}$ is:

$$\mathcal{K}(\mathbf{R}_{\mathbf{z}}) = \frac{1}{2} \left[Trace\left(\mathbf{R}_{\mathbf{z}}\right) - \log\det\left(\mathbf{R}_{\mathbf{z}}\right) - m \right]. \tag{3.18}$$

A whitening matrix is obtained when $\mathcal{K}(\mathbf{R}_{\mathbf{z}}) = 0$. This can be achieved by using the following update rule [23]:

$$\mathbf{V}_{k+1} = \mathbf{V}_k - \eta_k \left[\mathbf{z}_k \mathbf{z}_k^T - \mathbf{I} \right] \mathbf{V}_k, \tag{3.19}$$

where η_k is a variable adaptation step size.

3.3 Tracking changes in signal subspace

An important issue in adaptive subspace tracking algorithms is the ability to track possible changes in the dimension of the signal subspace. Eigenvalue inspection of $\mathrm{E}\left[\mathbf{y}\mathbf{y}^T\right]$ solves the problem of identifying the dimension in the case when $\mathbf{R}_{\mathbf{y}}$ is time invariant. However, in adaptive PCA the problem becomes more challenging since at each time k we receive a new observation vector $\mathbf{y}_k$. Thus, assuming that we know the number of principal eigenvalues at time step $k-1$ and based on the new information vector received at time k, we must decide if the number of sources has changed or remained the same.

A solution to this problem was proposed by Real et. al. [111]. They first consider the observation matrix $\mathbf{Y}_k = [\mathbf{y}_{k-L_w}, \ldots, \mathbf{y}_k]$ given by a window of length L_w. As a new information vector $\mathbf{y}_k$ becomes available, the energy $\mathcal{E}_0$ of the matrix $\mathbf{Y}_k$ is computed as the square of the Frobenius norm:

$$\| \mathbf{Y}_k \|_F^2 = \| \mathbf{Y}_{k-1} \|_F^2 - \| \mathbf{y}_{k-L_w} \|_2^2 + \| \mathbf{y}_k \|_2^2 . \tag{3.20}$$

From the matrix energy $\mathcal{E}_0$ we subtract successively larger sums of squares of the largest of the estimated eigenvalues λ_i until the difference lies below a chosen threshold value. For $i = 1, \ldots, m$ we compute:

$$\mathcal{E}_i = \mathcal{E}_0 - \sum_{l=1}^{i} \lambda_l^2 \tag{3.21}$$

and each value of $\mathcal{E}_i$ (including $\mathcal{E}_0$) is compared to the threshold. The number of times $\mathcal{E}_i$ exceeds the threshold is the estimated dimension of the signal subspace. However, some problems still arise from the selection of the threshold value. This can be set either theoretically from knowledge or assumption about the power in the orthogonal subspace or heuristically from estimates of that power [111].

In blind source separation it is also of great interest to detect changes in the mixing system. Paper IV proposed computing a sample covariance matrix in a relatively small processing sliding window and comparing the matrix to the covariance matrix constructed by the subspace tracker (3.17). The sample covariance matrix at time k in a window of N samples may be recursively computed by:

$$\mathbf{R}_k = \mathbf{R}_{k-1} + \frac{1}{N}\mathbf{y}_k\mathbf{y}_k^T - \frac{1}{N}\mathbf{y}_{k-N}\mathbf{y}_{k-N}^T. \tag{3.22}$$

A matrix of correlation coefficients is formed from the covariance matrices $\hat{\mathbf{R}}_\mathbf{y}$ and $\hat{\mathbf{R}}_k$. The following dimensionless expression

$$\frac{\| \hat{\boldsymbol{\mathcal{C}}}_\mathbf{y} - \hat{\boldsymbol{\mathcal{C}}}_k \|_F}{\| \hat{\boldsymbol{\mathcal{C}}}_\mathbf{y} \|_F} \tag{3.23}$$

is compared to a threshold value. If the value of (3.23) exceeds the threshold, the weighting used in the subspace tracking is reset to the initial weights so that recent measurements are weighted more heavily.

Whitening is an important stage for one class of separation algorithms. Hence, accurate values resulting from eigendecomposition are needed in order to obtain robust whitening. In

time-varying scenarios it is important to track both signal and noise subspaces. The method proposed in [70] performs well in slowly time-varying scenarios. A very good presentation on various techniques for subspace tracking is presented in [32].

Chapter 4

Adaptive Blind Source Separation Algorithms

4.1 Introduction

Blind separation algorithms may be categorized as batch or as on-line (real-time) algorithms based on the availability and treatment of the data. Batch algorithms operate on a separate set of multiple observations during each processing cycle. If update of the separation matrix is implemented by iterating over the whole block of data, the update is said to be adaptive. On-line algorithms update an existing separation matrix when a new information vector becomes available for processing, rather than determining an entirely new separation matrix. We want to emphasize the difference in adaptation between the two classes of algorithms.

In the rest of the chapter we examine adaptive on-line algorithms. Our attention is focused on several adaptive blind separation techniques. We start by introducing the concept of equivariance and we present the equivariant adaptive separation via independence (EASI) algorithm. Next we present nonlinear PCA class of algorithms and we continue with the application of state-variable models, Kalman filters, and particle filters to blind source separation. We then briefly show how the FIR-MIMO model can be converted to I-MIMO by using fractional sampling. The chapter ends with a discussion.

4.2 Equivariant algorithms

In the family of adaptive blind source separation algorithms there are algorithms whose behavior is independent of the mixing system. This property is called *equivariance* [23]. Examples of equivariant algorithms are the natural gradient algorithm [138] and the EASI algorithm [23]. In the following we will consider the EASI algorithm in detail.

An equivariant approach was first employed in a batch algorithm and was later extended to adaptive algorithms [23]. Let us assume that the number of sources is equal to the number of sensors, $m = n$. For the time being we consider the mixing model presented in (2.1):

$$\mathbf{y}_k = \mathbf{A}\mathbf{s}_k. \tag{4.1}$$

In batch processing it is assumed that all the observations $\mathbf{Y}_k = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$ are available to the receiver. From the definition of BSS, a blind estimator of the mixing matrix $\mathbf{A}$ is a function of $\mathbf{Y}_k$ only. This may be written as follows:

$$\hat{\mathbf{A}} = \mathcal{A}\left(\mathbf{Y}_k\right). \tag{4.2}$$

According to the mixing model (4.1), it was observed [23] that by multiplying the data by some matrix $\mathbf{G}$ has the same effect as multiplying $\mathbf{A}$ itself with $\mathbf{G}$, $\mathbf{G}\left(\mathbf{Y}_k\right) = \mathbf{G}\left(\mathbf{A}\mathbf{S}_k\right) = \left(\mathbf{G}\mathbf{A}\right)\mathbf{S}_k$, where $\mathbf{S}_k = [\mathbf{s}_1, \ldots, \mathbf{s}_k]$. An estimator $\mathcal{A}$ is said to be *equivariant* if for any invertible $m \times m$ matrix $\mathbf{G}$ satisfies:

$$\mathcal{A}\left(\mathbf{G}\mathbf{X}_k\right) = \mathbf{G}\mathcal{A}\left(\mathbf{X}_k\right), \tag{4.3}$$

where $\mathbf{X}_k = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$. A very important property of equivariant batch estimators is *uniform performance*. Let us assume that the sources are estimated as $\hat{\mathbf{s}}_k = \hat{\mathbf{A}}^{-1}\mathbf{y}_k$ where $\hat{\mathbf{A}}^{-1}$ is obtained from an equivariant estimator. Using (4.1), (4.2) and the equivariance property (4.3) we have:

$$\hat{\mathbf{s}}_k = \left[\mathcal{A}\left(\mathbf{Y}_k\right)\right]^{-1}\mathbf{y}_k = \left[\mathcal{A}\left(\mathbf{A}\mathbf{S}_k\right)\right]^{-1}\mathbf{A}\mathbf{s}_k = \left[\mathbf{A}\mathcal{A}\left(\mathbf{S}_k\right)\right]^{-1}\mathbf{A}\mathbf{s}_k = \mathcal{A}\left(\mathbf{S}_k\right)^{-1}\mathbf{s}_k. \tag{4.4}$$

The main result from the above derivation is:

$$\hat{\mathbf{s}}_k = \mathcal{A}\left(\mathbf{S}_k\right)^{-1}\mathbf{s}_k, \tag{4.5}$$

which may be interpreted as follows. The estimated sources signals $\hat{\mathbf{s}}_k$ using an equivariant estimator $\mathcal{A}$ for a particular realization $\mathbf{S}_k$ do not depend on the mixing matrix $\mathbf{A}$.

The adaptive algorithm developed by Cardoso [23] is based on *serial update*. In [23], a serial update algorithm is first defined as follows:

$$\mathbf{W}_{k+1} = \left[\mathbf{I} - \mu_k \mathcal{G}_{EASI}(\mathbf{x})\right] \mathbf{W}_k, \tag{4.6}$$

where $\mathcal{G}_{EASI}$ is an $m \times m$ matrix-valued function and μ_k is a sequence of adaptation steps. The system $\mathbf{W}_k$ is serially updated using left multiplication by the matrix $\mathbf{I} - \mu_k \mathcal{G}_{EASI}(\mathbf{x})$. This update technique exhibits uniform performance.

The EASI algorithm was derived by factoring the separating matrix as $\mathbf{W} = \mathbf{UV}$ where $\mathbf{V}$ is an $m \times n$ whitening matrix and $\mathbf{U}$ is an $m \times m$ orthogonal matrix. The goal [23] was to adaptively update the whitening matrix $\mathbf{V}$ and the orthogonal matrix $\mathbf{U}$ and then to combine them into a unique serial update rule in order to get a separating matrix $\mathbf{W}$.

The update of the whitening matrix is described in (3.19). The adaptation of the orthogonal matrix $\mathbf{U}$ proceeds via minimization of the following objective function:

$$\mathcal{J}(\mathbf{W}) = \mathrm{E}\left[f\left(\mathbf{x}\right)\right], \tag{4.7}$$

where $\mathbf{x} = \mathbf{UVs}$, $f\left(\mathbf{x}\right) = \sum_{i=1,m} |x_i|^4$. The following [23] update rule is obtained:

$$\mathbf{U}_{k+1} = \mathbf{U}_k - \mu_k \left[f'\left(\mathbf{x}_k\right) \mathbf{x}_k^T - \mathbf{x}_k f'\left(\mathbf{x}_k\right)\right] \mathbf{U}_k, \tag{4.8}$$

where $f'\left(\mathbf{x}_k\right)$ is the gradient of f at $\mathbf{x}_k$. By combining the updates for $\mathbf{V}$ and $\mathbf{U}$ we obtain the one-stage solution:

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu_k \left[\mathbf{x}_k \mathbf{x}_k^T - \mathbf{I} + g\left(\mathbf{x}_k\right) \mathbf{x}_k^T - \mathbf{x}_k g\left(\mathbf{x}_k\right)^T\right] \mathbf{W}_k, \tag{4.9}$$

where $g\left(\mathbf{x}\right) = \left[g_1(x_1), \ldots, g_m(x_m)\right]$ are m non-linear functions and the term $\mathbf{x}_k \mathbf{x}_k^T - \mathbf{I}$ has the effect of driving the diagonal elements of $\mathbf{WA}$ to unity. Recall that $\mathbf{W}$ is the separation matrix, $\mathbf{A}$ is the mixing matrix. Thus the product $\mathbf{WA}$ is actually a permutation matrix with arbitrary unknown scaling.

In order to preserve uniform performance an *ad hoc* [23] stabilization solution was proposed:

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu_k \left[\frac{\mathbf{x}_k \mathbf{x}_k^T - \mathbf{I}}{1 + \mu_k \mathbf{x}_k^T \mathbf{x}_k} + \frac{g\left(\mathbf{x}_k\right) \mathbf{x}_k^T - \mathbf{x}_k g\left(\mathbf{x}_k\right)^T}{1 + \mu_k \left|\mathbf{x}_k^T g\left(\mathbf{x}_k\right)\right|}\right] \mathbf{W}_k. \tag{4.10}$$

The denominator in (4.10) prevents the update term from taking large values. This may happen when $\mathbf{W}_k \mathbf{A}$ is very dissimilar to the identity matrix or when outliers are present in the received data.

4.3 Nonlinear PCA

Nonlinear PCA stems from the PCA learning rule [102]:

$$\mathbf{E}_{k+1} = \mathbf{E}_k + \mu_k \left[\mathbf{y}_k - \mathbf{E}_k \mathbf{x}_k\right] \mathbf{x}_k^T, \tag{4.11}$$

where $\mathbf{x}_k = \mathbf{E}_k^T \mathbf{y}_k$. The weight vectors $\mathbf{e}_k(i)$, $\mathbf{E}_k = [\mathbf{e}_k(1), \ldots, \mathbf{e}_k(m)]$, become orthonormal and tend to an m-dimensional principal eigenvector subspace of the correlation matrix $\mathrm{E}\left[\mathbf{y}_k \mathbf{y}_k^T\right]$. The nonlinear PCA learning rule replaces the output $\mathbf{x}_k$ with a nonlinear transform $g\left(\mathbf{x}_k\right) = g\left(\mathbf{W}_k^T \mathbf{y}_k\right)$. The learning rule (4.11) becomes:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k \left[\mathbf{y}_k - \mathbf{W}_k^T g\left(\mathbf{x}_k\right)\right] g\left(\mathbf{x}_k^T\right). \tag{4.12}$$

There is proof [73] that for the nonlinear PCA class of algorithms suitable nonlinearities are odd polynomial functions in the case of positive kurtotic sources and hyperbolic tangents in the case of negative kurtotic sources. When first introduced in [104], the update rule (4.12) was not motivated by an optimization criteria. However, it was shown in [72] that the update (4.12) is an approximate stochastic gradient algorithm minimizing the mean square error:

$$J(\mathbf{W}) = \mathrm{E}\left[\| \mathbf{y} - \mathbf{W} g\left(\mathbf{W}^T \mathbf{y}\right) \|^2\right]. \tag{4.13}$$

One of the very important conclusions in [103] is that if the effect of second order statistics is removed by whitening, the remaining higher order statistics can be used in nonlinear PCA to find the independent components. Thus, the cost function to be minimized is:

$$J(\mathbf{W}) = \mathrm{E}\left[\| \mathbf{z} - \mathbf{W} g\left(\mathbf{W}^T \mathbf{z}\right) \|^2\right]. \tag{4.14}$$

This leads to the following update rule:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k \left[\mathbf{z}_k - \mathbf{W}_k^T g\left(\mathbf{x}_k\right)\right] g\left(\mathbf{x}_k^T\right), \tag{4.15}$$

where $\mathbf{z}$ is the whitened data and $\mathbf{x}_k = \mathbf{W}_k^T \mathbf{z}_k$. No constraints on the matrix $\mathbf{W}$ are imposed for this learning rule. However, $\mathbf{W}$ is an orthogonal matrix, $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, for a suitable nonlinear function g if all the sources have the same distribution. We denote the orthogonal separation

matrix with $\mathbf{W}^o$. Under the orthogonality constraint we obtain:

$$\begin{aligned} J(\mathbf{W}^o) &= \mathrm{E}\left[\| \mathbf{z} - \mathbf{W}^o g\left[(\mathbf{W}^o)^T \mathbf{z}\right] \|^2\right] \\ &= \mathrm{E}\left[\| (\mathbf{W}^o)^T \mathbf{z} - (\mathbf{W}^o)^T (\mathbf{W}^o)\, g\left((\mathbf{W}^o)^T \mathbf{z}\right) \|^2\right] \\ &= \mathrm{E}\left[\| \mathbf{x} - g(\mathbf{x}) \|^2\right]. \end{aligned} \tag{4.16}$$

Choosing $g_i(\mathbf{z})$ as the odd quadratic function:

$$g_i(x) = \begin{cases} x^2 + x, & \text{if } x \geq 0 \\ -x^2 + x, & \text{if } x < 0 \end{cases} \tag{4.17}$$

the criterion (4.16) becomes

$$J(\mathbf{W}^o) = \sum_{i=1}^{m} \mathrm{E}\left[\left(x_i - x_i \pm x_i^2\right)^2\right] = \sum_{i=1}^{m} \mathrm{E}\left[x_i^4\right]. \tag{4.18}$$

Minimizing the nonlinear PCA criterion is equivalent to minimizing the sum of kurtosis of x_i.

Another nonlinear PCA solution based on approximate Recursive Least Squares (RLS) techniques was proposed in [75]. RLS algorithm converges faster than stochastic gradient algorithms. This makes RLS techniques feasible solutions to adaptive separation at the expense of higher complexity. The algorithm is inspired by the PAST algorithm introduced in [136] and described in section (3.2.1). In applying the separation algorithm, the data vector must be prewhitened. The RLS separation algorithm [75] can be summarized as follows:

$$\begin{aligned} \mathbf{q}_k &= \mathbf{g}\left((\mathbf{W}^o)_{k-1}^T \mathbf{z}_k\right) \\ \mathbf{p}_k &= \mathbf{z}_k - (\mathbf{W}^o)_{k-1}^T \mathbf{q}_k \\ \mathbf{k}_k &= \frac{\mathbf{P}_{k-1}\mathbf{q}_k}{\gamma + \mathbf{q}_k^T \mathbf{P}_{k-1}\mathbf{q}_k} \\ \mathbf{P}_k &= \frac{1}{\gamma}\mathrm{Tri}\left[\mathbf{P}_{k-1} - \mathbf{k}_k \mathbf{q}_k^T \mathbf{P}_{k-1}^T\right] \\ \mathbf{W}_k^o &= \mathbf{W}_{k-1}^o + \mathbf{p}_k \mathbf{k}_k^T, \end{aligned} \tag{4.19}$$

where $\mathbf{z}_k$ is the whitened input, $\mathbf{W}^o$ is the orthogonal separating matrix and the other variables are used in the update of the algorithm. The function g denotes an odd nonlinear function, a typical choice being $g(\cdot) = tanh(\cdot)$. This kind of adaptive update can be used for tracking if the statistics of the data or the mixing model vary slowly with time [74]. Following the PASTd extension, the algorithm can be modified to sequentially compute the weight vectors using a deflation technique [75].

4.4 State-variable model in BSS

The use of state-variable models in BSS was proposed by Salam [114] and has been investigated by other researchers as well [42, 46, 80, 140]. In each case the model is constructed differently. Both the initial sources or mixing matrix can be modeled as the state vector. Using a state-space model leads to the applicability of various types of Kalman filters for estimating states. Moreover, when using extended state-space models, the natural gradient method can be used for updating some of the model matrices [140]. A comprehensive review of linear Gaussian models is given in [113].

If we consider $\bar{\mathbf{a}}_k = vec(\mathbf{A}_k)$ to be the $nm \times 1$-dimensional vector obtained by stacking the columns of $\mathbf{A}$, then we have the following general state-space model [15, 53]:

$$\bar{\mathbf{a}}_{k+1} = \boldsymbol{\Upsilon}\left(\bar{\mathbf{a}}_k, \mathbf{w}_k\right) \tag{4.20}$$

$$\mathbf{y}_k = \boldsymbol{\Psi}\left(\bar{\mathbf{a}}_k, \mathbf{v}_k\right), \tag{4.21}$$

where $\bar{\mathbf{a}}_k$ represents the state vector, $\mathbf{y}_k$ is the observation vector, $\mathbf{w}_k$ is the state noise vector, $\mathbf{v}_k$ is the measurement noise vector, and $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are non-linear functions. Both noise sequences are considered to be zero-mean and white. The available information at step k is the sequence of observations $\mathbf{Y}_k = \{\mathbf{y}_i, i = 1, \ldots, k\}$ and the *pdf* of the state from the previous step $k-1$, $p\left(\bar{\mathbf{a}}_{k-1}|\mathbf{Y}_{k-1}\right)$. The task is to construct the *pdf* of the current state $p\left(\bar{\mathbf{a}}_k|\mathbf{Y}_k\right)$ based on the available information. This can be done recursively by using a prediction-correction type of update. Considering that $p\left(\bar{\mathbf{a}}_{k-1}|\mathbf{Y}_{k-1}\right)$ is known and taking into account the Markov property that $\bar{\mathbf{a}}_k$ depends only on $\bar{\mathbf{a}}_{k-1}$, the prediction of the state at time k is:

$$p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1}) = \int p(\bar{\mathbf{a}}_k|\bar{\mathbf{a}}_{k-1}) p(\bar{\mathbf{a}}_{k-1}|\mathbf{Y}_{k-1}) d\bar{\mathbf{a}}_{k-1}. \tag{4.22}$$

Then, at time step k the measurement $\mathbf{y}_k$ becomes available and can be used to update the prediction made in (4.22) via Bayes rule:

$$p(\bar{\mathbf{a}}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\bar{\mathbf{a}}_k) p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{k-1})}, \tag{4.23}$$

where the normalizing denominator is given by:

$$p(\mathbf{y}_k|\mathbf{Y}_{k-1}) = \int p(\mathbf{y}_k|\bar{\mathbf{a}}_k) p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1}) d\bar{\mathbf{a}}_k. \tag{4.24}$$

The recurrence relations (4.22) and (4.23) constitute a formal solution to the Bayesian recursive estimation problem.

Assume that Υ and Ψ are linear, the prior and posterior densities are Gaussian, and that $\mathbf{w}_k$ and $\mathbf{v}_k$ represent mutually uncorrelated, additive Gaussian noise. The model described by equations (4.20) and (4.21) is:

$$\bar{\mathbf{a}}_{k+1} = \mathbf{F}_k \bar{\mathbf{a}}_k + \mathbf{w}_k \tag{4.25}$$

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{s}_k + \mathbf{v}_k, \tag{4.26}$$

where $\mathbf{s}_k$ are the initial sources, $\bar{\mathbf{a}}_k$ are the elements of the mixing matrix $\mathbf{A}_k$ stacked in a vector form, $\mathbf{y}_k$ are the observations and $\mathbf{F}_k$ is the state transition matrix. The noise $\mathbf{w}_k$ is the state noise and $\mathbf{v}_k$ is the observational noise, both being considered to be Gaussian with covariance matrices $\mathbf{R}_\mathbf{w}$ and $\mathbf{R}_\mathbf{v}$. In the absence of *a priori* information, $\mathbf{F}_k$ is considered to be an identity matrix.

These two previous equations constitute the state-space model used by a Kalman filter. This model does not cover all situations. For instance, $\mathbf{v}_k$ and $\mathbf{w}_k$ may be correlated or noise may be colored. However, if the Gaussian assumption is employed, if we know $\mathbf{F}$ and $\mathbf{A}$ and if the covariance of $\mathbf{v}_k$ and $\mathbf{w}_k$ is known, then the filter is optimal. In other cases the filter is the best linear estimator. The goal of the Kalman filter is to find the minimum mean-square estimate of the state $\bar{\mathbf{a}}_k$. This is done by minimizing the trace of the filtered state-error covariance matrix $\mathrm{E}\left[\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T\right]$, where $\boldsymbol{\epsilon}_k$ is defined as $\boldsymbol{\epsilon}_k = \bar{\mathbf{a}}_k - \hat{\bar{\mathbf{a}}}_{k|k}$. This means that the Kalman filter is the linear minimum variance estimator of the state vector $\bar{\mathbf{a}}_k$ [55].

4.4.1 Particle filters

Everson [42] considered the problem of blind source separation with non-stationary mixing of stationary sources. The dynamics of the mixing system are modeled by a first order Markov process. The elements of the mixing matrix are the states and the goal is to determine the *pdf* of the state given the observations.

The idea of a particle filter is to approximate the posterior distribution of the state (the filtering density) with a set of possible state realizations or particles. Each particle is assigned a weight. The filtering density is approximated by a discrete distribution whose support is the set of

particles, with the probability mass of each particle being proportional to its weight. This is also known as *sequential importance sampling* (SIS) algorithm [15] (see [4] and references therein). As the number of samples becomes very large the SIS filter approaches the optimal Bayesian estimate. The particle filter algorithm specifies how the particles and their weights are propagated through time, to model the dynamics of the state and take into account the information in the new observations. Particle filters may be also considered as a practical solution to the model presented in (4.20)-(4.21). In [42], the model presented in equations (4.25)-(4.26) was used.

The source densities are modeled as generalized exponentials:

$$p(s^i|\check{r}_i,\check{\beta}_i,\check{\mu}_i) = \frac{\check{r}_i}{2\check{\beta}_i\Gamma(1/\check{r}_i)}\exp\left\{-\left|\frac{s^i-\check{\mu}_i}{\check{\beta}_i}\right|^{\check{r}_i}\right\}, \tag{4.27}$$

where the parameters $\check{\theta}_i = \left\{\check{r}_i,\check{\mu}_i,\check{\beta}_i\right\}$ must be calculated for each x^i at every stage of learning.

The problem is to track $\bar{\mathbf{a}}_k$ and to learn $\check{\theta}$ as new observations $\mathbf{y}_k$ become available. This involves finding the *pdf* of the state $p(\bar{\mathbf{a}}_k|\mathbf{Y}_k)$, where $\mathbf{Y}_k$ denotes the sequence of observations $\{\mathbf{y}_1,\ldots,\mathbf{y}_k\}$. To achieve this, a prediction-correction type of update is used. Particle filters represent the state density $p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1})$ using a cluster of N_p particles, each with probability mass [43]. Each particle's probability mass is modified using the state and observation equations, after which a new independent sample is obtained from the posterior $p(\bar{\mathbf{a}}_k|\mathbf{Y}_k)$ before proceeding to the next prediction/observation step.

The prediction of the state at time k is given by equation (4.22) where the prior density $p(\bar{\mathbf{a}}_{k-1}|\mathbf{Y}_{k-1})$ is modeled as Gaussian. Predicting the density $p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1})$ may be regarded as an estimate of $\bar{\mathbf{a}}_k$ prior to the observation of $\mathbf{y}_k$. The prediction stage is implemented as follows. Selecting N_p samples $\{\mathbf{w}_k^1,\mathbf{w}_k^2,\ldots,\mathbf{w}_k^{N_p},\}$ from the state noise density $\mathcal{N}(0,\mathbf{R}_\mathbf{w})$, each particle is propagated through the state equation (4.25) to form a new swarm of particles $\{\bar{\mathbf{a}}_{k|k-1}^1,\bar{\mathbf{a}}_{k|k-1}^2,\ldots,\bar{\mathbf{a}}_{k|k-1}^{N_p},\}$, where

$$\bar{\mathbf{a}}_{k|k-1}^l = \mathbf{F}\bar{\mathbf{a}}_{k-1|k-1}^l + \mathbf{w}_{k-1}^l \tag{4.28}$$

and $l = 1,\ldots,N_p$. If the particles $\{\bar{\mathbf{a}}_{k-1}^l\}$ are independent samples from $p(\bar{\mathbf{a}}_{k-1}|\mathbf{y}_{k-1})$, then $\bar{\mathbf{a}}_{k|k-1}^l$ are independent samples from $p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1})$. Thus the prediction stage implements equation (4.22).

The prediction represented by the swarm of particles $\{\bar{\mathbf{a}}_{k|k-1}^{l}\}$ from (4.28) has to be corrected upon the arrival of the new information data vector $\mathbf{y}_k$. Each particle is weighted by the likelihood of the observation $\mathbf{y}_k$ being generated by the mixing matrix represented by $\{\bar{\mathbf{a}}_{k|k-1}^{n}\}$. The particle probability masses q_k^l are assigned according to:

$$q_k^l = \frac{p(\mathbf{y}_k|\bar{\mathbf{a}}_{k|k-1})}{\sum_{l=1}^{N_p} p(\mathbf{y}_k|\bar{\mathbf{a}}_{k|k-1}^{l})}. \tag{4.29}$$

This procedure can be regarded as a discrete approximation of equation (4.23) where the prior $p(\bar{\mathbf{a}}_k|\mathbf{Y}_{k-1})$ is approximated by the sample $p(\bar{\mathbf{a}}_k|\mathbf{Y}_k)$. Finally, a resampling must be performed. The particles $\bar{\mathbf{a}}_{k|k-1}^{l}$ and weights q_k^l define a discrete distribution approximating $p(\bar{\mathbf{a}}_k|\mathbf{y}_k)$. These are resampled with replacement N_p times to form an approximate sample from $p(\bar{\mathbf{a}}_k|\mathbf{y}_k)$, the particles having equal weights. This sample is used for the next prediction.

After each update of the state density, the maximum *a posteriori* estimate of $\mathbf{A}_k$ is used to estimate the sources $\mathbf{s}_k$. Maximum-likelihood estimates of the source parameters $\check{\theta}_i = \{\check{r}_i, \check{\mu}_i, \check{\beta}_i\}$ are then determined from the sequences $\{s^i\}^k$.

4.4.2 Direct estimation of sources

An alternate state-space model for BSS is given in Paper IV. The algorithm is using a prewhitening stage. The sources are the states and are modeled as autoregressive (AR) processes. The state-space model used is the following:

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{G}\mathbf{w}_{k-1} \tag{4.30}$$

$$\mathbf{y}_k = \mathbf{A}_k\mathbf{x}_k + \mathbf{v}_k, \tag{4.31}$$

where the state noise weighting matrix $\mathbf{G}$ is assumed to be an identity matrix because the sources are statistically independent. In the context of BSS, both the transition matrix $\mathbf{F}_k$ and the mixing matrix $\mathbf{A}_k$ have to be estimated. Given a state-space model in which the noise sequences are white and Gaussian and in which the state and measurement transition matrices are linear, a Kalman filter [27] can be employed to estimate the state:

$$\hat{\mathbf{x}}_{k|k} = \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k\left[\mathbf{z}_k - \hat{\mathbf{A}}_k\mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right], \tag{4.32}$$

where $\mathbf{z}_k$ are the whitened observations and $\mathbf{K}_k$ is the Kalman gain. In the state estimation equation (4.32), an estimate of the mixing matrix $\mathbf{A}$ is also needed. The following update rule is used:

$$\hat{\mathbf{A}}_k = \hat{\mathbf{A}}_{k-1} + \tilde{\mathbf{y}}_k^{\mathbf{A}} \left(K_k^{\mathbf{A}}\right)^T, \tag{4.33}$$

where $\tilde{\mathbf{y}}_k^{\mathbf{A}} = \mathbf{z}_k - \hat{\mathbf{A}}_{k-1}\hat{\mathbf{x}}_k^{\mathbf{A}}$ is the innovation in estimating $\hat{\mathbf{A}}$ and $\hat{\mathbf{x}}_k^{\mathbf{A}} = g\left(\boldsymbol{\upsilon}_k\right)$, $\boldsymbol{\upsilon}_k = \hat{\mathbf{A}}^T\mathbf{z}_k$. The functions $g(\boldsymbol{\upsilon}_k) = [g(\upsilon_k^1), \ldots, g(\upsilon_k^m)]^T$ are non-linear score functions that can depend on the sign of each source's kurtosis. The gain $\mathbf{K}_k^{\mathbf{A}}$ used in estimating the mixing matrix is:

$$\mathbf{K}_k^{\mathbf{A}} = \frac{\mathbf{P}_{k|k-1}\hat{\mathbf{x}}_k^{\mathbf{A}}}{\left(\hat{\mathbf{x}}_k^{\mathbf{A}}\right)^T \mathbf{P}_{k|k-1}\hat{\mathbf{x}}_k^{\mathbf{A}} + 1}, \tag{4.34}$$

where $\mathbf{P}_{k|k-1}$ is the prediction error covariance matrix used in Kalman filter updates.

The state transition matrix $\mathbf{F}_k$ describes how the state sequence evolves over time. It may contain a low-order AR model for improving prediction. Each component of the state variable model (4.30) evolves as follows:

$$\mathbf{x}_k^i = \check{\mathbf{F}}^i\mathbf{x}_{k-1}^i + \mathbf{w}_{k-1}^i, \text{ with } \check{\mathbf{F}}^i = \begin{pmatrix} f^{i,1} & \ldots & f^{i,r-1} & f^{i,r} \\ 1 & \ldots & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & \ldots & 1 & 0 \end{pmatrix} \tag{4.35}$$

where $f^{i,r}$ are the AR coefficients, r is the order of the AR model, $\mathbf{x}_k^i = \begin{bmatrix} x_k^i & x_{k-1}^i \cdots x_{k-r+1}^i \end{bmatrix}^T$ is a vector of past r states and $\mathbf{w}^i = [w_i \ 0 \cdots 0]^T$ has r elements. In $\mathbf{x}_k^i$ we are interested in only the predicted state x_k^i. This type of processing models how states evolve over time and allows for noise attenuation. The AR coefficients may be recursively estimated in many ways. A low complexity method is obtained by using RLS. If the source signals do not exhibit an autoregressive structure, the matrix $\mathbf{F}_k$ is considered to be an identity matrix.

4.4.3 General state-space models for separation/deconvolution

Zhang and Cichocki proposed very general mixing and demixing models covering both linear [140, 141] and nonlinear [29] systems. A general framework of state-space approaches for multichannel blind deconvolution of both linear and nonlinear system was presented in [142]. In the following, we will refer only to the linear mixing and demixing systems.

The linear state mixing model is described by a state equation of the form:

$$\check{\mathbf{u}}_{k+1} = \check{\mathbf{A}}\check{\mathbf{u}}_k + \check{\mathbf{B}}\mathbf{s}_k + \check{\mathbf{L}}\boldsymbol{\xi}_k, \tag{4.36}$$

where $\check{\mathbf{u}}$ is the $r \times 1$ vector of the system, $\mathbf{s}$ is the $m \times 1$ vector of input signals, $\check{\mathbf{A}}$ is the $r \times r$ state mixing matrix, $\check{\mathbf{B}}$ is the $r \times m$ input mixing matrix and $\boldsymbol{\xi}$ is the process noise. The n-dimensional vector of sensor signals is a linear combination of the states and inputs in the form:

$$\mathbf{y}_k = \check{\mathbf{C}}\check{\mathbf{u}}_k + \check{\mathbf{D}}\mathbf{s}_k + \boldsymbol{\theta}_k, \tag{4.37}$$

where $\check{\mathbf{C}}$ is the $n \times r$ output mixing matrix, $\check{\mathbf{D}}$ is the $n \times m$ input-output mixing matrix and $\boldsymbol{\theta}$ is the sensor noise of the mixing system.

Considering the mixing model from equations (4.36)-(4.37), the separation task is to recover the original sources from observation $\mathbf{y}$ without prior knowledge of the source signals and the state-space matrices $\left[\check{\mathbf{A}}, \check{\mathbf{B}}, \check{\mathbf{C}}, \check{\mathbf{D}}\right]$. In [140, 141, 142] it is proposed that the demixing model is another linear state-space system described as follows:

$$\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k + \mathbf{B}\mathbf{y}_k + \mathbf{L}\boldsymbol{\xi}_R \tag{4.38}$$

$$\mathbf{x}_k = \mathbf{C}\mathbf{u}_k + \mathbf{D}\mathbf{y}_k, \tag{4.39}$$

where the input $\mathbf{y}$ of the demixing model is the output of the mixing model and $\boldsymbol{\xi}_R$ is the reference model noise. The matrices $\boldsymbol{\mathcal{W}} = [\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]$ are the parameters to be determined in learning.

When the matrices $\check{\mathbf{A}}, \check{\mathbf{B}}, \check{\mathbf{C}}$ in the mixing model and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in the demixing model are null matrices, the problem is simplified to a standard ICA problem. The state-space equation of the mixing model reduces to:

$$\mathbf{y}_k = \check{\mathbf{D}}\mathbf{s}_k + \boldsymbol{\theta}_k, \tag{4.40}$$

where $\check{\mathbf{D}}$ is in fact the mixing matrix from equation (2.1). The separation model is simplified to:

$$\mathbf{x}_k = \mathbf{D}\mathbf{y}_k, \tag{4.41}$$

where $\mathbf{D}$ is separation matrix $\mathbf{W}$ from equation (2.5). Under the demixing model, the remaining task is to estimate the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ in order to achieve separation of the sources.

Two methods were proposed for updating the set of matrices $\boldsymbol{\mathcal{W}}$. In the first approach [140] the matrices $\mathbf{A}$ and $\mathbf{B}$ are assumed to be known and they are fixed during learning. Matrix $\mathbf{D}$ is

also restricted to being an $m \times m$ nonsingular matrix. Following Amari's derivation for natural gradient methods [9], Cichocki and Zhang used natural gradient algorithm in updating the matrix $\mathbf{D}$ and standard gradient method for updating $\mathbf{C}$:

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \eta\varphi\left(\mathbf{x}_k\right)\mathbf{u}_k^T \tag{4.42}$$

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \eta\left[\mathbf{I} - \varphi\left(\mathbf{x}_k\right)\mathbf{y}_k^T\mathbf{D}_k^T\right]\mathbf{D}_k, \tag{4.43}$$

where η is a learning rate and $\varphi(\mathbf{x})$ is a vector of nonlinear activation functions. Typically, if a source signal is super-Gaussian, one can choose $\varphi(x) = tanh(x)$. In the case of sub-Gaussian sources $\varphi(x) = x^3$ can be used [8, 39]. From equations (4.42) and (4.43) we note that the natural gradient algorithm [11] is covered as a special case of the learning algorithm for linear state demixing model. In the case when $\check{\mathbf{A}}, \check{\mathbf{B}}, \check{\mathbf{C}}$ are null matrices and also $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are considered to be zero, then the learning algorithm (4.43) is the same as the natural gradient learning algorithm [11].

In another approach [141] the blind deconvolution problem is divided into separation and state estimation. Recursive updates are proposed for the matrices $\mathbf{C}$ and $\mathbf{D}$. In order to compensate for the model bias and reduce the effect of noise, a Kalman filter is employed to estimate the state vector $\mathbf{u}_k$. For the objective function $l(\mathbf{x}, \mathbf{W})$, the natural gradient $\nabla l(\mathbf{x}, \mathbf{W})$ is the steepest ascent direction of the objective function. Following the same derivations as in their previous paper [140], Cichocki and Zhang introduced a new search direction [141]. This lead to the following updates of $\mathbf{C}$ and $\mathbf{D}$:

$$\mathbf{C}_{k+1} = \mathbf{C}_k + \eta\left[\left(\mathbf{I} - \varphi\left(\mathbf{x}_k\right)\mathbf{x}_k^T\right)\mathbf{C}_k - \varphi\left(\mathbf{x}_k\right)\mathbf{u}_k^T\right] \tag{4.44}$$

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \eta\left(\mathbf{I} - \varphi\left(\mathbf{x}_k\right)\mathbf{x}_k^T\right)\mathbf{D}_k. \tag{4.45}$$

Instead of adjusting the matrices $\mathbf{A}$ and $\mathbf{B}$ directly, in [141] it is proposed to estimate the state $\mathbf{u}_{k+1}$ using Kalman filter. The Kalman filter dynamics are given as follows:

$$\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k + \mathbf{B}\mathbf{y}_k + \mathbf{K}\mathbf{r}_k + \boldsymbol{\xi}_R, \tag{4.46}$$

where $\mathbf{K}$ is the Kalman gain and $\mathbf{r}_k$ is the innovation. Since updating matrices $\mathbf{C}$ and $\mathbf{D}$ will produce an innovation in each learning step, the notion of a hidden innovation was introduced

[141] as follows:

$$\mathbf{r}_k = \Delta\mathbf{x}_k = \Delta\mathbf{C}\mathbf{u}_k + \Delta\mathbf{D}\mathbf{y}_k, \tag{4.47}$$

where $\Delta\mathbf{C} = \mathbf{C}_{k+1} - \mathbf{C}_k$ and $\Delta\mathbf{D} = \mathbf{D}_{k+1} - \mathbf{D}_k$. The hidden innovation presents the adjusting direction of the output of the demixing system and is used to generate an *a posteriori* state estimate. Once we obtain the hidden innovation, the conventional Kalman filter [55] can be applied in order to estimate the state vector $\mathbf{u}_k$.

4.5 Application in blind equalization

In communications, systems with multiple receiver antennas and oversampling of received signals can be modeled as MIMO linear time invariant systems for which there are a variety of detection and estimation algorithms [57, 84, 106, 108]. In the presence of ISI, the received signals are linear mixtures not only of current independent symbols from different sources, but also of the adjacent symbols from the same sources. Using fractional sampling combined with source separation, it is in general possible to recover the original sources directly as long as there is enough information resulting from oversampling [137, 143].

In [137] an FIR-SIMO equalization problem was converted into a blind separation problem. An FIR-MIMO case was considered in [143]. The idea in this scheme is to transform an FIR-MIMO system into an I-MIMO system by fractional sampling. The channel mixing impulse response will be denoted by $\mathbf{f}_{ij}$ for the path from transmitter i to receiver j. For simplicity we assume that $L = \max(L_{ij})$ is the channel length. For illustration purposes and without loss of generality, we consider the case of a 2 input 2 output system (see [143] for a more detailed case). The noise-free model is:

$$\mathbf{x}(t) = \mathbf{T}(t)\mathbf{s}(k), \tag{4.48}$$

where $\mathbf{x}(t) = [x_1(t) \;\; x_2(t)]^T$ and $\mathbf{s}[k] = [s_1[k] \;\; s_2[k]]^T$. The channel mixing matrix has a structure $\mathbf{T} = [\bar{\mathbf{t}}_1(t) \;\; \bar{\mathbf{t}}_2(t)]^T$ with $\bar{\mathbf{t}}_1(t) = [\mathbf{t}_{11}(t) \;\; \mathbf{t}_{21}(t)]$ and $\bar{\mathbf{t}}_2(t) = [\mathbf{t}_{12}(t) \;\; \mathbf{t}_{22}(t)]$. Sampling at the rate M/T with an oversampling factor M, the jth receiver signal x_j at time $t = lT +$

$iT/M + \zeta_j$ with $i = 0, 1, \ldots, 2L+1$ gives:

$$x_j(lT + \frac{iT}{M} + \zeta_j) = \sum_{n=l-L}^{l+L+1} \mathbf{f}_j((l-n)T + \frac{iT}{M} + \zeta_j)\mathbf{s}[k] \tag{4.49}$$

$l = 0, 1, \cdots$, where ζ_j is some unknown sampling offset at the j-th receiver and T is the symbol period [143]. Since samples acquired within one symbol interval are used in BSS problems, we choose the oversampling factor $M \geq 2L + 2$. If the sampling offsets are zero the oversampling factor is $M = 2L + 1$. Let us use the following notation:

$$\mathbf{T}_i[l] \stackrel{\text{def}}{=} \left[\mathbf{t}_1(lT + \frac{iT}{M} + \zeta_1)^T \quad \mathbf{t}_2(lT + \frac{iT}{M} + \zeta_2)^T\right]^T \tag{4.50}$$

$$\mathbf{y}_i[l] \stackrel{\text{def}}{=} \left[x_1(lT + \frac{iT}{M} + \zeta_1) \quad x_2(lT + \frac{iT}{M} + \zeta_2)\right]^T \tag{4.51}$$

then

$$\begin{aligned} \mathbf{y}_j[l] &= \sum_{k=l-L}^{l+L+1} \mathbf{T}_j[l-n]\mathbf{s}[k] \\ &= \sum_{k=-L-1}^{L} \mathbf{T}_j[n]\mathbf{s}[l-k]. \end{aligned} \tag{4.52}$$

Based on the above derivation the following instantaneous mixing model can be built:

$$\begin{pmatrix} \mathbf{y}_0[l] \\ \vdots \\ \mathbf{y}_{2L+1}[l] \end{pmatrix} = \bar{\mathbf{A}} \cdot \begin{pmatrix} \mathbf{s}[l-L] \\ \vdots \\ \mathbf{s}[l+L+1] \end{pmatrix} \tag{4.53}$$

with

$$\bar{\mathbf{A}} = \begin{pmatrix} \mathbf{T}_0[L] & \cdots & \mathbf{T}_0[-L-1] \\ \vdots & & \vdots \\ \mathbf{T}_{2L+1}[L] & \cdots & \mathbf{T}_{2L+1}[-L-1] \end{pmatrix} \tag{4.54}$$

of dimension $2(2L+2) \times 2(2L+2)$. The matrix $\bar{\mathbf{A}}$ is a constant matrix and the model (4.53) represents the noise-free I-MIMO case. All the $2(2L+2)$ components are assumed to be independent and $\bar{\mathbf{A}}$ is of full rank. A more comprehensive derivation is presented in [143].

In paper III the previous fractional sampling scheme was used with a Kalman-based source separation algorithm in order to perform adaptive blind equalization. The case of slowly time-varying channels was also addressed. It is important to note that the complexity of the fractional sampling scheme increases with L, hence, this scheme is suitable only for channels with small memory. As an example, for a typical GSM channel having $L = 4$ taps the oversampling factor must be at least $M = 10$. Moreover, one should be careful in choosing the whitening algorithm since in this case the number of received signals is equal to the number of transmitted signals.

4.6 Discussion

BSS approaches using prewhitening usually perform well, but may suffer from a serious loss of accuracy when some of the source signals s are weak or if the mixing matrix $\mathbf{A}$ is ill-conditioned [23]. Equivariant algorithms avoid these problems by using an overall system matrix $\mathbf{C}_k = \mathbf{A}_k\mathbf{W}_k$ describing the mixing and demixing process which depends only on the previous $\mathbf{C}_{k-1}$ and on the output vector $\mathbf{x}_k = \mathbf{B}_k\mathbf{y}_k = \mathbf{C}_k\mathbf{s}_k$. However, if noise is present the equivariance property is lost.

It has been demonstrated [55] that the recursive least-squares algorithms perform better and exhibit faster convergence than stochastic gradient algorithms. This was also confirmed experimentaly in [76] when RLS-type of algorithm was applied to blind source separation. Due to its adaptive form, the RLS algorithm can be used for tracking statistics of the data or for tracking slow changes in the mixing model.

State-space models are a natural way of describing the BSS problem if one models the mixing matrix elements [42] or the sources as states, see paper IV. This model also attenuates the effects of noise. Another gain is that if the states are time-varying, tracking can be performed. An important aspect is also the information that is considered known about the sources. If, for example, the sources exhibit some autoregressive structure or other properties, these should be included in the model, as shown in Paper IV. This is more problematic in the case of information-bearing models. For example, in EEG there is no information about the mixing system, so the task of the separation algorithm is more difficult. In this situation, more general algorithms are merely

intended to provide adequate performance in situations for which additional useful system information is not available. Particle filters allow nonlinear state-space model and multimodal source distributions whereas the Kalman based source separation is more fitted to unimodal source distributions and linear mixing. An overall conclusion is that when using state-space models the state can contain either the elements of the mixing matrix or the sources.

In [140] and [141] state-space models for blind separation and deconvolution have been proposed. In [140] a state-space model is considered but the model is used only to compute the estimated sources. Gradient-based updates are used to compute parts of the model transition matrices while the other matrices are considered to be known and fixed during learning. In the second approach suggested in [141] a Kalman-based update is used for computing the state and gradient methods are used for computing the part of the model transition matrices. In [140], noise was not considered in the model for purposes of simplicity. This simplifies the derivations but does not allow for noise attenuation. Despite the fact that state-space models are considered, there is no direct implementation of Kalman filtering to solve the separation problem. Moreover, using gradient techniques to update part of the model transition matrices needed in the innovation update of the Kalman filter may lead to loosing the optimality of the Kalman state update.

The solution presented in Paper III tackles the problem of blind equalization via blind separation. Even if the algorithm performs well in slowly time-varying scenarios it may be impractical in real applications where the channels vary rapidly and may experience deep fades, case in which the separation algorithm will fail to perform. In fact the critical part is the subspace tracking method which cannot track fast time-varying eigenstructure. Moreover, very large oversampling factors may result making the algorithm not feasible in real applications.

In conclusion, there is no single algorithm that would be the best in all cases, hence a fair comparison between all algorithms is difficult to make since they are designed to provide a solution to a specific problem.

Chapter 5

Adaptive MIMO Channel Equalization

5.1 Introduction

Multi-path propagation in wireless channels results in distorted signals at the receiver. When propagation delays are longer than a symbol period, resulting inter-symbol interference (ISI) can cause high error rates unless removed. The process of removing ISI is known as equalization. Communications channels can be time selective, frequency selective or time-frequency selective. In the time selective case equalization involves estimating the time-varying amplitude and phase distortions of the channel and using this estimates to compensate their effects [109, 124]. In the frequency selective case, the estimated channel is used to adjust the parameters of a filter which compensates for frequency-dependent channel effects. The filter may be linear, such as that of a transversal equalizer, or nonlinear, such as those used in a decision feedback equalizer. Alternatively, the filter may implement a maximum likelihood sequence estimator [109, 124]. A recent comprehensive review of adaptive equalization techniques is given in [124]. There are also several books on the topic [55, 109, 110].

Equalization of time-varying FIR-MIMO channels is a very important research topic in communications. In general, the design of an optimal equalizer requires precise knowledge of channel parameter values [132]. Channel parameters are usually estimated using a limited number of data samples. From this perspective we can identify [132] three types of channel estimation approaches: training-based, blind and semi-blind. Assuming perfect knowledge of the MIMO

channel, a maximum likelihood sequence estimator (MLSE) is the optimum receiver, but has exponential complexity even when implemented using the Viterbi algorithm [109]. Another method is to bypass the channel estimation step and to directly compute parameter values for a desired equalizer structure [116]. Still another option is to estimate the channel and then to design an equalizer based on the estimated channel [14, 83, 117, 123, 128]. Such equalizers use *joint tracking and equalization* algorithms.

There are several classes of blind equalization algorithms. These classes include Bussgang methods, algorithms explicitly based on higher order statistics, joint maximum likelihood channel and data estimation algorithms, and algorithms based on cyclostationary second-order statistics [109]. Several technical journals have devoted recent issues to blind algorithms [1, 2]. There are also a number of books on blind methods [55, 57, 93, 109]. Techniques for blind equalization of single-input multiple-output (SIMO) and multiple-input multiple-output (MIMO) are presented in [47]. Despite the advantages of blind equalization, such as a gain in capacity, there are also major drawbacks. For example ambiguities always remain, such as the rotation of the constellation. Some of the methods may have poor convergence and some of the channels are not identifiable.

Semi-blind methods offer possible solutions for the problems of blind equalization techniques. Semi-blind techniques can exhibit the useful properties of both training-based and blind algorithms. In semi-blind techniques, the presence of a small number of training signals allows resolution of ambiguities related to mis-convergence and channel identifiability [47]. On the other hand, making use of statistical information from the non-training signals allows semi-blind techniques to outperform training-based methods that exploit only the known training signals [35]. In time-frequency selective channels there is a need for first estimating the channel and for then computing equalizer parameter values. Due to possible rapid time variations the channel parameters should be tracked continuously and equalizer parameters should be updated accordingly. In this chapter we consider semi-blind techniques.

The chapter begins with a description of the time-varying channel model used in simulations. The problem of channel estimation is then considered, with an emphasis on Kalman filter techniques. The problem of estimating the noise statistics needed in the recursive update of a Kalman

filter is also addressed. Several decision feedback equalizer (DFE) structures for MIMO channels are presented. The chapter ends with a discussion.

5.2 Channel characterization

Fundamentally, mobile radio communication channels are time-varying multipath channels. Since the performance of digital radio communication systems is strongly affected by multipath propagation in the form of scattering, reflection and diffraction, channel models are of great interest [18, 67, 79, 96, 100, 109, 110, 120, 125].

A time-varying radio channel (TVC) may be represented by a two-dimensional channel impulse response $h(t, \tau)$. See, for instance, Figure 5.1 for an example. Multipath propagation results in time dispersion of the transmitted signal, which is visible on the τ axis of $h(t, \tau)$. Time variations of the channel are given on the t axis of $h(t, \tau)$ where T is the symbol duration.

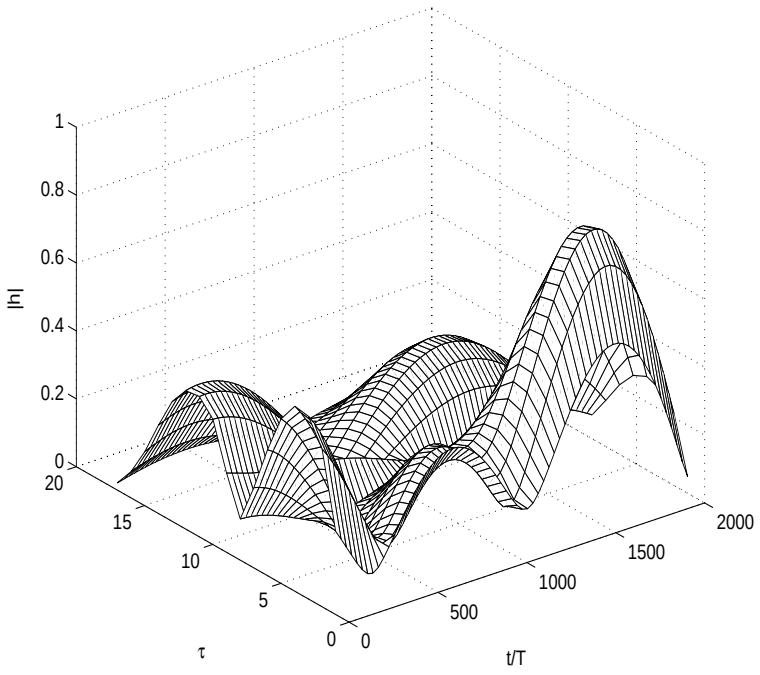

Figure 5.1: COST-207 'Hilly Terrain' channel, receiver speed of 90 km/h.

Two types of fading characterize mobile communications: large-scale fading, which is due to the motion over large areas and small-scale fading which is due to small changes in position [120]. Small-scale fading is often modeled using Rayleigh distribution when the multiple re-

flective paths are large in number and if there is no line-of-sight (LOS) signal component, the envelope of the received signal may be statistically described by a Rayleigh *pdf*. When there is a dominant non-fading component, such as a LOS propagation path, the small-scale fading envelope is modeled as a Ricean *pdf*. As the amplitude of the non-fading component approaches zero, the Ricean *pdf* approaches a Rayleigh *pdf*. The Ricean distribution is often described in terms of a parameter K. This is also known as the Ricean factor and completely specifies the Ricean distribution as it is defined as the ratio between the deterministic signal power and the variance of the multipath [110].

In wireless communications many physical factors in the radio propagation channel cause fading. Typically there is no LOS path between the mobile units and the base station. Consequently, the received signal consists of multiple copies of the transmitted signal that arrive at the receiver through different indirect paths. When LOS is present, the channel can be modeled as containing an LOS component and also as containing multipath components, as we will see later in this chapter. The randomly distributed amplitudes, phases and arrival angles of these multipath copies of the transmitted signal cause fluctuations in the received signal power, thereby introducing fading. In addition to (multipath) fading, multipath propagation also lengthens the time required for the main portion of the transmitted signal to reach the receiver. This phenomenon is quantified by maximum excess delay τ_{max}. In the case of a single transmitted signal waveform τ_{max} represents the time between the first and the last received component. Depending on the relative durations of the maximum excess delay and the symbol period, multipath fading is conventionally classified into either frequency-flat fading or frequency-selective fading [120]. Multipath fading is frequency-selective when the symbol period is smaller than the maximum excess delay. Thus the channel induces intersymbol interference (ISI). The fading is flat when all the received multipath components arrive within the symbol period. The coherence bandwidth f_0 is a measure of the range of frequencies over which the channel response can be considered "flat". In other words, coherence bandwidth is the range of frequencies over which two frequency components have a strong potential for amplitude correlation [110].

Other causes of fading are the frequency offsets between two sources. The signal from one source undergoes Doppler shift due to relative motion between a transmitter and a receiver. Also,

there may be a carrier frequency mismatch between transmit and receive oscillators. Frequency offsets result in frequency modulation on the transmitted signal, and thereafter cause channel time-variations that are quantified by coherence time. Depending on the relative values of the coherence time T_c and the symbol period, a fading channel can be categorized either as time-flat when the symbol period is much less than the channel coherence time. Otherwise, it is time-selective.

Frequency-selectivity and time-selectivity are two different properties of a fading channel. Taking into account combinations of time-selectivity and frequency-selectivity, fading channels are conventionally categorized into one of the following four types:

- Flat fading channels (channels are both time- and frequency-flat).
- Frequency-selective fading channels (channels are frequency-selective but time-flat).
- Time-selective fading channels (channels are time-selective but frequency-flat).
- Doubly-selective fading channels (channels are both frequency- and time-selective).

The wide sense stationary uncorrelated scattering (WSSUS) linear time-variant channel model is widely used to model signal propagation in mobile communications environment. The WSSUS model was introduced by Bello [18] and it was further investigated, for example in [60]. According to [60], the following model can be written:

$$h_{ij}(t,\tau) = \frac{1}{\sqrt{N_l}} \sum_{l=1}^{N_l} e^{j(2\pi f_{d,l} t + \theta_l)} h_{RF}(\tau - \tau_l), \tag{5.1}$$

where N_l is the number of echo paths, $f_{d,l}$ is the Doppler spread, θ_l is the angular spread and $h_{RF}(t)$ is the impulse response of the receive filter. For each delay τ, the channel is given by selecting:

1. N_l Doppler frequencies $f_{d,l}$ from a random variable with classical Jakes pdf in $(-f_{d,max}, f_{d,max})$. The maximum Doppler spread can be expressed like $f_{d,max} = v/\lambda_w$, where v is the mobile station speed and λ_w is the signal wave length.
2. N_l initial phases θ_l from a uniform distributed random variable in $[0, 2\pi]$.

3. N_l echo delay times τ_l. Each delay spread is a random variable with probability density function proportional to the mean power delay profile of the propagation environment.

The uncorrelated scattering assumption leads to an FIR channel model in which all the taps vary independently of one other. That is, the time-variations of the tap coefficients are mutually uncorrelated while exhibiting the same time-correlation behavior. The physical situation underlying this model is the existence of a few large scatterers far from the mobile receiver and the existence of a small number of scatterers in the vicinity of the mobile receiver. Classical analysis of digital transmission through a fading medium models $h_{ij}(t,\tau)$ as zero mean random variables. In certain applications, such as cellular communications, a direct non-fading path may also exist, superimposed on the fading path. In this case, the coefficients $h_{ij}(t,\tau)$ have non-zero mean (Rician fading). The overall non-zero mean channel is then [34, 128]:

$$\check{h}_{ij}(t,\tau) = \tilde{h}_{ij}(\tau) + h_{ij}(t,\tau), \tag{5.2}$$

where $\tilde{h}_{ij}(\tau)$ is a constant mean and $\mathrm{E}\left[h_{ij}(t,\tau)\right] = 0$.

As pointed out in [93], accurate mathematical channel models are based on collected measurements of actual channels. The COST-207 final report [100] defines propagation channels that appear in GSM systems. Measurements have been made over typical bandwidths of 10 to 20 MHz at or near 900 MHz. Four propagation environments are described in the project: Typical Urban (TU), Bad Urban (BU), Hilly Terrain (HT) and Rural Area (RA), each of them having specific parameter sets. These COST-207 propagation environments are determined by individual delay distributions that are piecewise exponential functions. A summary of COST-207 channel parameters is presented in Appendix 1. With these parameters defined, various COST-207 TVCs can be modeled. A MATLAB implementation of COST-207 model was presented in [19]. It is also important to mention that COST-207 played an important role in supporting the Group Spécial Mobiles in their work that lead to the original design of the Global System for Mobile Communications (GSM).

Several other COST [95] projects have been active in the area of channel modeling. The major goals of COST 259 [98] was devoted to an area of huge expansion in telecommunications, that of high-rate wireless data transmission. Three main areas were addressed, namely radio

systems, antennas and propagation and network aspects. Final models and algorithms have been achieved and many results have been obtained including techniques for OFDM transmission, enhancement of TDMA systems, and near-far resistant techniques in CDMA. Also directional radio channels for which indoor/outdoor measurements have been taken were characterized and adaptive antennas for GSM and wideband CDMA were analyzed.

Projects like COST 273 [101] are curently active. The main objective of this action is to increase knowledge of the radio aspects of mobile broadband multimedia networks, by exploring and developing new methods, models, techniques, strategies and tools to further the implementation of fourth generation mobile communication systems.

Other COST projects (COST 227, COST 231 & COST 280) have the goal of defining feasible systems for mobile communications based on integration of satellite and terrestrial networks [97, 101, 99]. In these projects one of the tasks is deriving a channel model for Earth-satellite and terrestrial paths above about 20 GHz.

Classical channel models provide information on signal power level distribution, Doppler shifts of received signals. Modern spatial channel models incorporate concepts as time delay spread, Angle Of Arrival or adaptive antenna geometries [88, 106]. An overview of spatial channel models is presented in [41], research on channel models has been done also at AT&T Labs [135], several other models are presented in special issues of technical journals, such the one in [125].

5.3 Recursive Channel Estimation

Training-based channel estimation consists of classical techniques that estimate the channel from a known training sequence and observed channel outputs. The mode of operation is “train-before-transmit”[132], which is effective when the channel does not have significant time variations. In most cases, training methods appear as robust methods but present some disadvantages. For example, the effective data rate decreases as a non-negligible portion of a data packet is devoted to training symbols [35]. In GSM, about 20% of the bits in a burst are used for training.

Blind equalization techniques [79] allow the estimation/equalization of the channel or the

equalizer based only on the received data without any training symbols. In other words, the channel estimation is performed while information symbols are being transmitted, hence, it is a "train-while-transmit"-type of transmission [132]. The major advantage of this type of transmission is the improved effective data rate. One drawback is that blind channel estimates converge slowly from random initial tap values. In rapidly time-varying channel conditions we need to have some information about the system (e.g. a small number of transmitted symbols) at least for initialization purposes. In the case of deep fades, the channel tracking algorithm may fail. Thus we need some new information in order to restore tracking after the fade. In blind methods, some ambiguities always remain (e.g. rotation of the constellation pattern), and some channels may be unidentifiable (for example if subchannels in SIMO model have common zeros).

Perhaps one of the best definitions of the notion of *semi-blind* channel estimation/equalization is given by referring to the use of the known information. Training sequence (TS) methods base the parameter estimation only on the received signal containing known symbols. All of the other observations, which may contain unknown symbols, are ignored [35]. Blind methods are based on all the received data and on knowledge of the structural and statistical properties of the transmitted data, but not on explicit knowledge of input symbols. For example, it may be known that the symbols have the property of constant modulus or are i.i.d.

The purpose of semi-blind methods is to combine the benefits of both training sequence-based and blind methods. Semi-blind techniques, due to the fact that they incorporate the information of known symbols, avoid the possible pitfalls of blind methods and with only a few known symbols, any channel, SISO or MIMO, becomes identifiable [35, 85]. Semi-blind techniques appear very interesting from a performance point of view, as their performance can be superior to that of TS or blind techniques separately. Semi-blind techniques may be applicable in cases in which TS and blind methods fail individually [35]. This may be the case of fast TVC when TS and blind methods cannot estimate and track the channel, especially when deep fades occur, but semi-blind techniques can offer a viable solution.

Different choices are available for implementing channel estimation and equalization, depending on the channel modeling and on the complexity allowed for each task. Recursive algorithms are needed in order to perform real-time computation and to avoid recomputing everything

at the arrival of new symbol. From the family of recursive algorithms, Kalman filter algorithms offer the best results in terms of channel estimation and tracking under the assumption of states which obey the Gauss-Markov model, the linearity of the model and known noise statistics. In highly demanding scenarios (i.e. high speed of the receiver) Kalman filter algorithms offer a good tradeoff between complexity and performance. They have been successfully applied to the problem of channel estimation in [66, 82, 128] and in papers V and VI.

5.3.1 Kalman filter for channel tracking

Regular Kalman filters have been applied in SISO and MIMO channel estimation [81, 82, 119]. The state-space model for MIMO channels may be written as:

$$\begin{aligned} \mathbf{h}(k) &= \mathbf{F}\mathbf{h}(k-1) + \mathbf{w}(k-1) \\ \mathbf{y}(k) &= \bar{\mathbf{X}}(k)\mathbf{h}(k) + \mathbf{v}(k), \end{aligned} \tag{5.3}$$

where $\bar{\mathbf{X}}$ is a $n \times nmL$ data matrix defined as:

$$\bar{\mathbf{X}}(k) = [x_1(k)\mathbf{I}_m \dots x_n(k)\mathbf{I}_m \dots x_1(k-L+1)\mathbf{I}_m \dots x_n(k-L+1)\mathbf{I}_m], \tag{5.4}$$

$\mathbf{I}_m$ is an $m \times m$ identity matrix and the channel taps are stacked in a vector of length nmL:

$$\begin{aligned} \mathbf{h}(k) = \big[&h_{11}^0(k) \dots h_{1n}^0(k) \dots h_{m1}^0(k) \dots h_{mn}^0(k) \dots \\ &h_{11}^{L-1}(k) \dots h_{1n}^{L-1}(k) \dots h_{m1}^{L-1}(k) \dots h_{mn}^{L-1}(k)\big]^T. \end{aligned} \tag{5.5}$$

Matrix $\mathbf{F}$ is the state transition matrix. The condition for applying a conventional Kalman filter to the model of (5.3) is that the state variables $\mathbf{h}(k)$ are Gauss-Markov random processes. This is nearly satisfied in many practical applications for which the channel can be described as Rayleigh with uncorrelated scattering [66]. Another condition is that the observations are linear, which is the case of the model from equation (5.3). A Kalman filter [55] can be then applied to the model of (5.3), leading to an estimate of the state.

5.3.2 Modeling the channel as an AR process

The time correlation of the channel can be exploited in order to allow better channel estimates. A low-order AR model can be used to describe channel time evolution [14, 66, 128]. An extended

Kalman filter (EKF) has been employed in a spread-spectrum SISO environment [66]. The state vector consisted of the code delay, the Doppler spread, and the tap coefficients. The tap coefficients were assumed to be uncorrelated. EKF has been employed due to the nonlinear nature of the measurement equation. The channel was modeled as a first-order AR model and an extension to order-p AR models was also presented. However, the derivation and analysis of the algorithm were for the first-order AR case and no method for estimating the state transition matrix (which contained the AR parameters) was given. The most important fact is that the channel was considered to be WSSUS, thus only time-correlation of the individual channels was exploited by the AR model.

The uncorrelated scattering assumption was removed in [128, 129], where the problem of estimating fading channel environments with correlated coefficients in a SISO case was considered. The key part of the work in [14, 128, 129] was modeling the channel as a multichannel AR process of order r. This leads to the following expression for the state equation:

$$\mathbf{h}_k = \sum_{l=1}^{r} \mathbf{F}_l \mathbf{h}_{k-l} + \mathbf{w}_k, \tag{5.6}$$

where $\mathbf{h}_k = \begin{bmatrix} h_{k,0} & \cdots & h_{k,L} \end{bmatrix}^T$ are the channel taps, $\mathbf{w}_k$ is a i.i.d. circular complex Gaussian vector and the matrix $\{\mathbf{F}_l\}_{l=1}^{r}$ is a $L \times L$ matrix containing the unknown model coefficients. In order to simplify the application of a Kalman filter, the following state-space model may to be built:

$$\bar{\mathbf{h}}_{k+1} = \bar{\mathbf{F}}_k \bar{\mathbf{h}}_k + \mathbf{J}\mathbf{w}_k \tag{5.7}$$

$$y_k = \bar{\mathbf{x}}_l^T \bar{\mathbf{h}}_k + v_k, \tag{5.8}$$

where $\bar{\mathbf{h}}_k = \begin{bmatrix} \mathbf{h}_k^T & \cdots & \mathbf{h}_{k-r+1}^T \end{bmatrix}^T$, $\mathbf{w}_k = \begin{bmatrix} w_k & 0 & \cdots & 0 \end{bmatrix}^T$ and $\mathbf{J} = [\mathbf{I} \ 0 \ \cdots \ 0]$. The data vector is defined as $\bar{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \end{bmatrix}^T$ where $\mathbf{x}_k = \begin{bmatrix} x_k & x_{k-1} & \cdots & x_{k-L} \end{bmatrix}^T$. The state matrix $\bar{\mathbf{F}}$ is:

$$\bar{\mathbf{F}} = \begin{pmatrix} \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_r \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ & \ddots & \ddots & \vdots \\ \mathbf{0} & & \mathbf{I} & \mathbf{0} \end{pmatrix}.$$

Given the data y_k and $\bar{\mathbf{x}}$, [128] shows how to estimate $\bar{\mathbf{F}}_k$, which contains the AR parameters. It is argued that in multichannel time-series the AR parameters are uniquely identified by the cor-

relation matrices $\mathbf{R}_{hh}(\tau) = \mathrm{E}\left[\bar{\mathbf{h}}_k \bar{\mathbf{h}}_{k+\tau}^H\right]$. By post multiplying (5.6) by $\bar{\mathbf{h}}_{k-\tau}^H$ and taking expected value of both sides, it is found that:

$$\mathbf{R}_{hh}^H(\tau) = \sum_{l=1}^{p} \mathbf{F}_l \mathbf{R}_{hh}^H(\tau - l) + \sigma_w^2 \delta(\tau)\mathbf{I}, \quad \tau = 0, \ \cdots \ , r. \tag{5.9}$$

By solving equation (5.9), the parameter matrices $\mathbf{F}_l$ are obtained. In order to save computation, an adaptive gradient method to solve (5.9) and to update $\hat{\mathbf{F}}_l$ can be used [128]. However, since $\bar{\mathbf{h}}$ is not directly observed, it follows that solving (5.9) is not a trivial task. In [128] two methods are proposed for estimating $\mathbf{R}_{hh}(\tau)$. One is based on Least Squares (LS) and the other on higher order statistics (HOS). The parameters of interest are the estimates of the (i, j) element of the matrix $\mathbf{R}_{hh}(\tau)$ for $\tau = 0, \ \cdots \ , p$:

$$[\mathbf{R}_{hh}(\tau)]_{i,j} = r_h(\tau; i-1, j-1), \quad i, j = 1, \ \cdots \ , L+1 \tag{5.10}$$

$$r_h(\tau; i, j) \stackrel{\text{def}}{=} \mathrm{E}\left[h_{k,i} h_{k+\tau,j}^*\right]. \tag{5.11}$$

Using the notation $\boldsymbol{\theta}_{\tau,true} \stackrel{\text{def}}{=} [r_h(\tau;0,0) \ \cdots \ r_h(\tau;0,L) \ r_h(\tau;1,0) \ \cdots \ r_h(\tau;L,L)]$, it is shown [128] that the LS solution of the linear regression

$$y_k y_{k+\tau}^* = \boldsymbol{\phi}_{k,\tau}^H \boldsymbol{\theta}_\tau + \sigma_v^2 \delta_\tau + e_{k,\tau} \tag{5.12}$$

for $k = 0, \cdots, N-1$ will yield an unbiased estimate of $\boldsymbol{\theta}_{\tau,true}$ under some very general assumptions, where $\boldsymbol{\phi}_{k,\tau}^H = [x(k)x^*(k+\tau), \ldots, x(k)x^*(k+\tau-L), x(k-1)x^*(k+\tau), \ldots, x(k-L)x^*(k+\tau-L)]$. In (5.12), the estimation error is given by $e_{k,\tau} = y_k y_{k+\tau}^* - \mathrm{E}\left[y_k y_{k+\tau}^* | \bar{\mathbf{x}}_k\right]$.

Another method for estimating the lags $r_h(\tau, q, q+\tau)$ employs HOS. Closed-form expressions for the tap correlations can be derived using fourth order statistics:

$$\hat{r}_h(\tau, q_0, q_1) = \begin{cases} \sigma_x^{-4} \mathrm{C}_{y_2 x_2}(\tau; \tau - q_1, -q_0) & q_1 \neq q_0 + \tau \\ \left(\gamma_{4x} + \sigma_x^4\right)^{-1} \mathrm{C}_{y_2 x_2}(\tau; -q_0, -q_0) & q_1 = q_0 + \tau \end{cases}, \tag{5.13}$$

where $\gamma_{4x} = cum\left[x_k, x_k, x_k^*, x_k^*\right]$.

The paper [128] addresses the problem when a LOS component exits. It follows that a mean component $\tilde{\mathbf{h}}$ is present according to equation (5.2). Combining (5.2) and (5.8) the following model is derived:

$$\bar{y}_k = \bar{\mathbf{x}}_l^T \check{\mathbf{h}}_k + y_k, \tag{5.14}$$

where $\check{\mathbf{h}}_k = \tilde{\mathbf{h}} + \bar{\mathbf{h}}_k$. The constants $\tilde{\mathbf{h}}$ can be consistently estimated via LS solution of equation (5.14), and y_k can be recovered. In a more general case, the mean can be modeled as varying slowly, so that a first order AR process can be used [34] to describe its time evolution. By combining this with the AR process of order r used for the channel taps, a new state-space formulation can be written [34].

5.3.3 Estimating noise statistics

As discussed above, the Kalman filter is widely used for channel estimation. Typically, only assumptions about the linearity of the state-space model and the Gauss-Markov structure of the state are invoked when speaking about the optimality of the Kalman filter. Very important components are assumed known, including the statistics of the state and measurement noises. Optimal Kalman filtering requires accurate values of these statistics. Hence, in communications it is of great importance to

estimate these values using the received measurements and possibly the transmitted symbols.

The problem of identification of noise covariances was addressed in [92] where a batch method was proposed. This method was applied in [118] in the context of SISO equalization of time-varying channels. It was further extended to an on-line MIMO case in Paper VII.

The noise estimation algorithm has two stages: covariance estimation and testing for the whiteness of the innovations. The noise statistics computation is based on a covariance matching method [92]. The measurement noise covariance matrix is found using the theoretical and estimated covariances of the innovation sequence. By matching these two expressions of the innovation covariance and writing the remaining terms in a recursive manner, a recursive formula for computing the measurement noise covariance matrix can be found. See Paper VII for details. The same approach is used for finding a recursive formula for the state noise covariance matrix. A residual process is defined and by matching its theoretical and sample covariance matrix a recursive formula for the state noise covariance is derived.

Adaptive methods for estimating the noise statistics were also introduced in [13, 69]. The method in [13] was based on a Bayesian type of approach and the method derived in [69] on a correlation matching method. Both methods found the same adaptive update for the observation

noise covariance matrix but different formulas for the state noise covariance. A comprehensive review of parameter estimation techniques related to Kalman filter is given in [91].

A necessary and sufficient condition for a Kalman filter to operate optimally is that innovations sequence is white (zero mean, uncorrelated). A recursive non-parametric method [87, 122] for testing the whiteness of the innovation process has been applied in paper VII. In the case of MIMO channels, components of an innovation vector are assumed to be mutually uncorrelated. Hence, whiteness of each component is tested individually. New sequences are formed from the innovation processes taking the sign of their samples [122]. A run is defined as a set of identical symbols contained between two different symbols. A sequence would be considered non-random if there are either too many or too few runs and random otherwise.

5.4 Decision Feedback Equalization in MIMO systems

Decision Feedback Equalizer (DFE) has received much attention recently in the channel equalization community. This is due to the fact that DFEs structure offer intersymbol interference (ISI) cancelation with reduced noise enhancement [25, 26]. More of the research has been concentrated on the SISO case [25, 26, 61] but there is also research in the MIMO case [6, 139]. A DFE is a nonlinear equalizer that employs previously detected symbols on the current symbol to be detected. The use of the previously detected symbols makes the equalizer output a nonlinear function of the data. Due to the nonlinear feedback nature of DFE, symbol errors introduced by noise may trigger bursts of errors which can lead to poor symbol error rate (SER) performance. Because the DFE uses past source symbol estimates to generate new decisions, finding the desired parameterization from a cold start, where no information about transmitted symbols is available, is a difficult task [25].

Let us consider y_k to be a received signal at time k. The structure of a SISO DFE is presented in Figure 5.2. The task of the DFE is to estimate the transmitted symbols $\hat{x}_k$. In the case of an FIR-DFE, equalization is achieved via feedforward $\mathbf{f} = \left(f_0, \ldots, f_{N_f}\right)^T$ and feedback $\mathbf{d} = (d_1, \ldots, d_{N_d})^T$ filters yielding soft estimates:

$$z(k) = \sum_{q=1}^{N_f} f_q y_{k-q} - \sum_{q=1}^{N_d} d_q \hat{x}_{k-q} \tag{5.15}$$

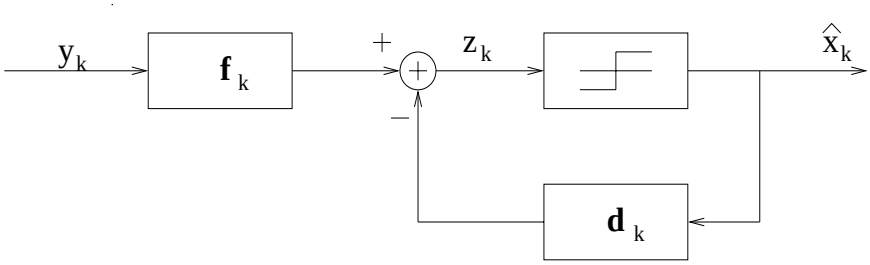

Figure 5.2: SISO DFE structure.

under the assumption of correct past decisions. The length of the feedforward (FF) filter is N_f and the length of the feedback (FB) filter is N_d. A DFE with a FF filter of length N_f and a FB filter of length N_d will be written as DFE(N_f,N_b). The symbol estimate $\hat{x}_k$ at time k is obtained by:

$$\hat{x}(k) \quad = \quad \arg\min_{\alpha \in \mathcal{X}} |\alpha - z(k)|, \tag{5.16}$$

where $\mathcal{X}$ is a finite alphabet.

The way in which the FF and FB filters are computed can be used to separate DFEs into two classes, one class that uses the channel in order to find the filters at each time step, for example minimum

mean square error DFE (MMSE-DFE) [14, 25] and a second class in which the filters are computed adaptively using a stochastic gradient descent algorithm [25] or other recursive techniques [61]. The first approach has the advantage that it is potentially more robust to the channel time-variations [45], however, it depends on good channel estimates. It has been also shown that in certain conditions this type of DFE can be less complex than adaptively updating the DFE coefficients [117]. The second class is more prone to error propagation since the error will be propagated in time as opposed to the first class where at each time step we estimate from the data the new filters. Moreover, it has been shown [25] that stochastic gradient descent DFEs suffer from ill convergence if they are not initialized near the MMSE-DFE solution.

5.4.1 MIMO MMSE-DFE

Let us start with the definition of the MIMO model used in communications. We will refer to the model from Figure 2.1. In order to employ the same notation that is commonly used in equalization community we denote the channel taps using h instead of a and the transmitted symbols using x instead of s. The rest of the notation remains the same as for the I-MIMO case. The noise is denoted by v and the received signals by y. The following model is considered:

$$\mathbf{y}_k = \sum_{l=0}^{L} \mathbf{H}_l \mathbf{x}_{k-l} + \mathbf{v}_k, \tag{5.17}$$

where $\mathbf{H}_l$ is the $n \times m$ lth MIMO channel matrix, $\mathbf{x}_{k-l}$ is a $m \times 1$ input vector at time $k-l$ and L is the maximum length of all of the mn channel impulse response, i.e. $L = \max_{i,j} L^{(i,j)}$.

A very useful classification was made in [121] where the following terminology was used. If there are no connections between the filters of different channels it is said to be a *non-connected* (NC)-FF/NC-FB a MIMO DFE. If there are connections between the FF and FB filters of each channel it is said to be a *fully-connected* (FC)-FF/FC-FB a MIMO DFE.

SIMO MMSE-DFE

In this subsection we present an important result which links the SISO case to the MIMO one, namely the Single Input Multiple Output (SIMO) model. This type of MMSE-DFE was introduced in [25]. The structure of the system is depicted in Figure 5.3. The model used is a special case of (5.17) with only one antenna at the transmitter. The received signal at antenna j is:

$$y_k^j = \left(\mathbf{h}_k^{1j}\right)^T \mathbf{x}_k + v_k^j, \tag{5.18}$$

where $\mathbf{h}_k^{1j}$ is a $L \times 1$ vector containing L channel taps of the path $1j$ at time k, $\mathbf{x}_k$ contains L past symbols sent by the transmit antenna and v_k^j is the AWGN at antenna j. The received vector at time k has the form $\mathbf{y}_k = [y_1 \cdots y_n]^T$. The SIMO MMSE-DFE giving soft decisions $\hat{z}_k$ has the structure:

$$\begin{aligned} \hat{z}_k &= \bar{\mathbf{f}}^T \bar{\mathbf{y}} - \mathbf{d}\hat{\mathbf{x}} \\ &= \sum_{j=1}^{n} \left(\sum_{q=0}^{N_f} f_q^{(j)} y_{k-q}^{(j)} \right) - \sum_{q=1}^{N_d} d_q \hat{x}_{k-q}, \end{aligned} \tag{5.19}$$

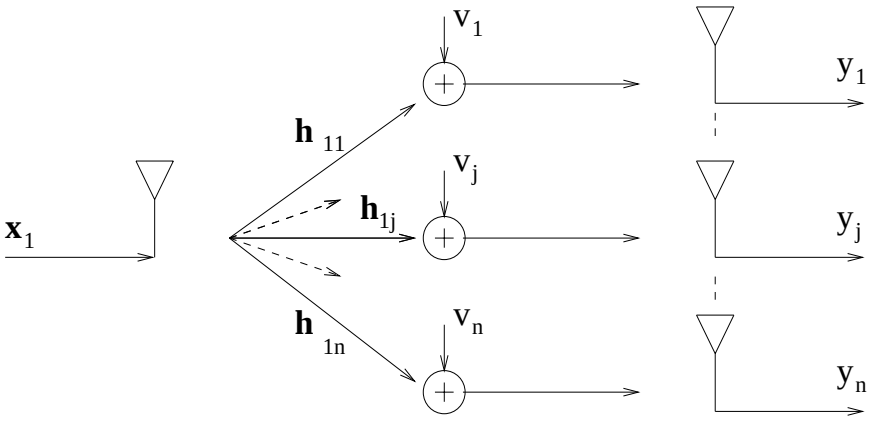

Figure 5.3: SIMO model.

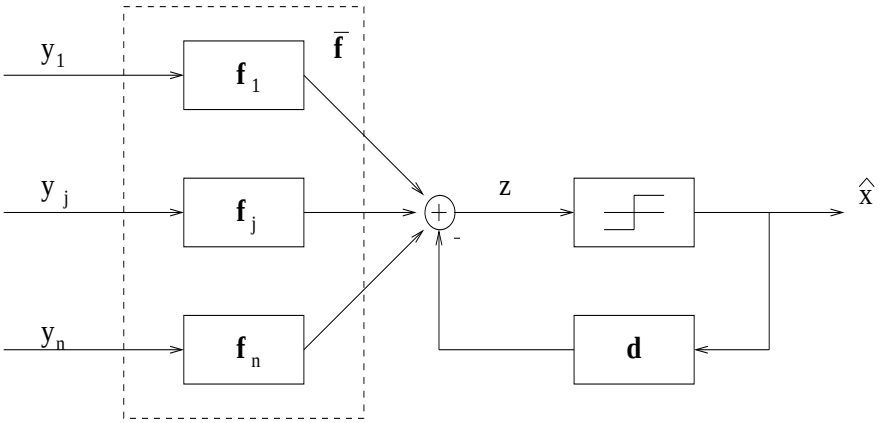

Figure 5.4: SIMO MMSE-DFE.

where $\bar{\mathbf{y}}_k = \left[\mathbf{y}_1^T \cdots \mathbf{y}_j^T \cdots \mathbf{y}_n^T\right]^T$, with $\mathbf{y}_k^j = \left[y_k^j \;\; y_{k-1}^j \cdots \; y_{k-N_f}^j\right]^T$. In other words, $\bar{\mathbf{f}}$ is a $nN_f \times 1$ vector which contains FF filters that are applied to the inputs $\bar{\mathbf{y}}_k$ and has the form $\bar{\mathbf{f}} = \left[\mathbf{f}_1^T \cdots \mathbf{f}_n^T\right]^T$. The detected symbols $\hat{x}_k$ are found by using equation (5.16).

The optimum FF and FB filters are estimated by minimizing the cost function:

$$\mathcal{J} = \mathrm{E}\{(x_{k-\Delta} - \hat{z}_k)^2\} \tag{5.20}$$

with respect to $\bar{\mathbf{f}}$ and $\mathbf{d}$. The following expressions are found [25]:

$$\bar{\mathbf{f}} = \left(\bar{\mathbf{H}}^H \mathbf{P} \bar{\mathbf{H}} + \sigma_v^2 \mathbf{I}\right)^{-1} \bar{\mathbf{H}}^H \mathbf{e}_\Delta \tag{5.21}$$

$$\mathbf{d} = \mathbf{M}^T \bar{\mathbf{H}} \bar{\mathbf{f}}, \tag{5.22}$$

where $\mathbf{P} = \mathbf{I} - \mathbf{M}\mathbf{M}^T$, $\mathbf{e}_\Delta = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is a standard basis vector, with one at

the position Δ, $0 \leq \Delta \leq N_{ch}$ and $N_{ch} = L + N_f$. The $N_{ch} + 1 \times nN_f$ matrix $\bar{\mathbf{H}}$ is build as $\bar{\mathbf{H}} = [\mathbf{H}_{11} \cdots \mathbf{H}_{1j}]$, the channel convolution matrices $\mathbf{H}_{1j}$ of dimension $N_{ch}+1 \times N_f$, are defined as:

$$\mathbf{H}_{1j}(k) = \begin{pmatrix} \hat{h}_{1j(0;k)} & 0 & \dots & 0 \\ \hat{h}_{1j(1;k)} & \hat{h}_{1j(0;k)} & \ddots & \vdots \\ \hat{h}_{1j(2;k)} & \hat{h}_{1j(1;k)} & & 0 \\ \vdots & \hat{h}_{1j(2;k)} & & \hat{h}_{1j(0;k)} \\ \hat{h}_{1j(L-1;k)} & \vdots & & \hat{h}_{1j(1;k)} \\ 0 & \hat{h}_{1j(L-1;k)} & & \hat{h}_{1j(2;k)} \\ \vdots & 0 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \hat{h}_{1j(L-1;k)} \end{pmatrix}, \text{ and } \mathbf{M} = \begin{pmatrix} 0_{(\Delta+1)\times N_d} \\ \mathbf{I}_{N_d \times N_d} \\ 0_{(N_{ch}-N_d-\Delta)\times N_d} \end{pmatrix}. \tag{5.23}$$

Fully connected MIMO MMSE-DFE

Different techniques for FC MIMO MMSE-DFE can be found in the literature. In [139] minimization of the geometric MSE (defined as the determinant of the symbol estimation error covariance matrix) was used to find the FF and FB filters. A general derivation for finite length MMSE-DFE was introduced in [5] for a SISO case and in [6] for a MIMO case. In [5], the problem of fractionally spaced FF filter and colored source and noise was considered. By making the assumption of a sufficiently long feedback filter, i.e. $N_d > N_f$, the problems of delay optimization and statistics of the recovery error, i.e. the error between the transmitted and estimated symbol, are addressed. In [6] the FF and FB filters are restricted to be FIR, the assumption of equal numbers of inputs and outputs is relaxed and a parallel structure that allows faster computation is presented. For simplicity, a 2×2 model of the MIMO MMSE-DFE used in [6] is presented in Figure 5.5. We will refer to this DFE as FC MIMO MMSE-DFE. This is due to the existence of the cross-FF filters and cross-FB filters.

The model introduced in (5.17) was used with the lth MIMO channel matrix $\mathbf{H}_l$ of dimension $\check{l}n \times m$. The parameter $\check{l}$ is the sampling coefficient. Over a block of N_f symbol periods (5.17)

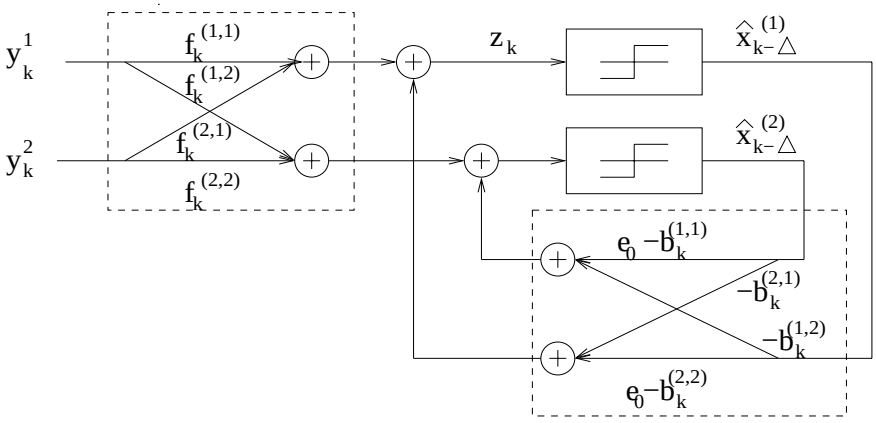

Figure 5.5: FC MIMO MMSE-DFE structure.

can be rewritten as:

$$\bar{\mathbf{y}}_{k+N_f-1:k} = \bar{\mathbf{H}}\bar{\mathbf{x}}_{k+N_f-1:k-L} + \bar{\mathbf{v}}_{k+N_f-1:k}, \tag{5.24}$$

where $\bar{\mathbf{H}}$ is a $\check{l}n(N_f-1) \times m(N_f+L-1)$ block matrix having the structure:

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_L & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & & \ddots & & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_L \end{bmatrix} \tag{5.25}$$

and $\bar{\mathbf{y}}_{k+N_f-1:k} = \begin{bmatrix} \mathbf{y}_{k+N_f-1}^T & \mathbf{y}_{k+N_f-2}^T & \cdots & \mathbf{y}_k^T \end{bmatrix}^T$, $\bar{\mathbf{x}}_{k+N_f-1:k-L} = \begin{bmatrix} \mathbf{x}_{k+N_f-1}^T & \mathbf{x}_{k+N_f-2}^T & \cdots & \mathbf{x}_{k-L}^T \end{bmatrix}^T$, $\bar{\mathbf{v}}_{k+N_f-1:k} = \begin{bmatrix} \mathbf{v}_{k+N_f-1}^T & \mathbf{v}_{k+N_f-2}^T & \cdots & \mathbf{v}_k^T \end{bmatrix}^T$.

Defining the $n(N_f+L) \times n(N_f+L)$ input autocorrelation matrix

$$\mathbf{R}_{xx} \stackrel{\text{def}}{=} \mathrm{E}\left[\bar{\mathbf{x}}_{k+N_f-1:k-L}\bar{\mathbf{x}}_{k+N_f-1:k-L}^*\right] \tag{5.26}$$

and the $(m\tilde{l}N_f) \times (m\tilde{l}N_f)$ noise autocorrelation matrix

$$\mathbf{R}_{vv} \stackrel{\text{def}}{=} \mathrm{E}\left[\bar{\mathbf{v}}_{k+N_f-1:k}\bar{\mathbf{v}}_{k+N_f-1:k}^*\right] \tag{5.27}$$

the input-output cross-correlation and the output auto-correlation matrices are given by:

$$\mathbf{R}_{xy} \stackrel{\text{def}}{=} \mathrm{E}\left[\bar{\mathbf{x}}_{k+N_f-1:k-L}\bar{\mathbf{y}}_{k+N_f-1:k}^*\right] = \mathbf{R}_{xx}\bar{\mathbf{H}}^* \tag{5.28}$$

$$\mathbf{R}_{yy} \stackrel{\text{def}}{=} \mathrm{E}\left[\bar{\mathbf{y}}_{k+N_f-1:k}\bar{\mathbf{y}}_{k+N_f-1:k}^*\right] = \bar{\mathbf{H}}\mathbf{R}_{xx}\bar{\mathbf{H}}^* + \mathbf{R}_{vv}. \tag{5.29}$$

The FIR MIMO MMSE-DFE consists of an FF filter matrix

$$\bar{\mathbf{F}}^* \stackrel{\text{def}}{=} \left[\mathbf{F}_0^* \;\; \mathbf{F}_1^* \;\; \cdots \;\; \mathbf{F}_{N_f-1}^*\right] \tag{5.30}$$

with N_f matrix taps $\mathbf{F}_i$, each of size $(m\check{l} \times n)$, and an FB filter matrix equal to

$$[\mathbf{I}_n \;\; 0_{n\times nN_d}] - \mathbf{B}^* \stackrel{\text{def}}{=} \left[(\mathbf{I}_m - \mathbf{B}_0^*) \;\; -\mathbf{B}_1^* \;\; \cdots \;\; -\mathbf{B}_{N_b}^*\right]. \tag{5.31}$$

The following matrix is defined [6]:

$$\mathbf{B}^* = \left[\mathbf{B}_0^* \;\; \mathbf{B}_1^* \;\; \cdots \;\; \mathbf{B}_{N_d}^*\right] \tag{5.32}$$

containing $N_d + 1$ matrix taps $\mathbf{B}_i$, each of size $n \times n$. Furthermore, the following matrix is defined: $\bar{\mathbf{B}}^* \stackrel{\text{def}}{=} [\mathbf{0}_{n\times n\Delta} \;\; \mathbf{B}^*]$, where $0 \leq \Delta \leq N_f + L - 1$ is the decision delay that satisfies the condition $\Delta + N_d + 1 = N_f + 1$. The error vector of the MIMO MMSE-DFE at time k is given by:

$$\check{\mathbf{e}}_k = \bar{\mathbf{B}}^* \bar{\mathbf{x}}_{k+N_f-1:k-L} - \mathbf{F}^* \bar{\mathbf{y}}_{k+N_f-1:k}. \tag{5.33}$$

Applying the orthogonality principle, which states that $\mathrm{E}\left[\check{\mathbf{e}}_k \bar{\mathbf{y}}_{k+N_f-1:k}\right] = 0$, the optimum FF filter matrix is found [6]:

$$\bar{\mathbf{F}}_{opt}^* = \bar{\mathbf{B}}_{opt}^* \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1}. \tag{5.34}$$

Using the partitioning $\mathbf{R} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}$ where $\mathbf{R}_{11}$ is of size $n(\Delta+1) \times n(\Delta+1)$,

$$\bar{\mathbf{B}}_{opt} = \begin{bmatrix} \mathbf{I}_{m(\Delta+1)} \\ \mathbf{R}_{12}^* \mathbf{R}_{11}^{-1} \end{bmatrix} \mathbf{C}, \tag{5.35}$$

where $\mathbf{C}^* \stackrel{\text{def}}{=} [\mathbf{0}_{n\times n\Delta} \;\; \mathbf{I}_n]$ and $\mathbf{R} = \mathbf{R}_{xx}^{-1} + \bar{\mathbf{H}}^* \mathbf{R}_{vv}^{-1} \bar{\mathbf{H}}$.

Non-connected MIMO MMSE-DFE

In this subsection the MIMO MMSE-DFE derived in Papers V and VI is briefly presented. This type of DFE belongs to the NC category, meaning that pairs of feedforward-feedback filters for each received signal based on the MMSE criterion are computed. The derivation is based on the assumption that the input and noise processes are uncorrelated. This scheme has also lower computational complexity.

The model used is the one presented in (5.17). The channel convolution matrices are $\hat{\mathbf{H}}_{ij}$ of dimension $N_{ch} \times N_f$, where $N_{ch} = L + N_f - 1$ is defined as:

$$\hat{\mathbf{H}}_{ij}(k) = \begin{pmatrix} \hat{h}_{ij(0;k)} & 0 & \dots & 0 \\ \hat{h}_{ij(1;k)} & \hat{h}_{ij(0;k)} & \ddots & \vdots \\ \hat{h}_{ij(2;k)} & \hat{h}_{ij(1;k)} & & 0 \\ \vdots & \hat{h}_{ij(2;k)} & & \hat{h}_{ij(0;k)} \\ \hat{h}_{ij(L_h-1;k)} & \vdots & & \hat{h}_{ij(1;k)} \\ 0 & \hat{h}_{ij(L_h-1;k)} & & \hat{h}_{ij(2;k)} \\ \vdots & 0 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \hat{h}_{ij(L_h-1;k)} \end{pmatrix}. \quad (5.36)$$

In this case the structure of the NC MIMO MMSE-DFE is the one shown in Figure 5.6. Applying the feedforward filter to the past N_f received observations and the feedback filter to the

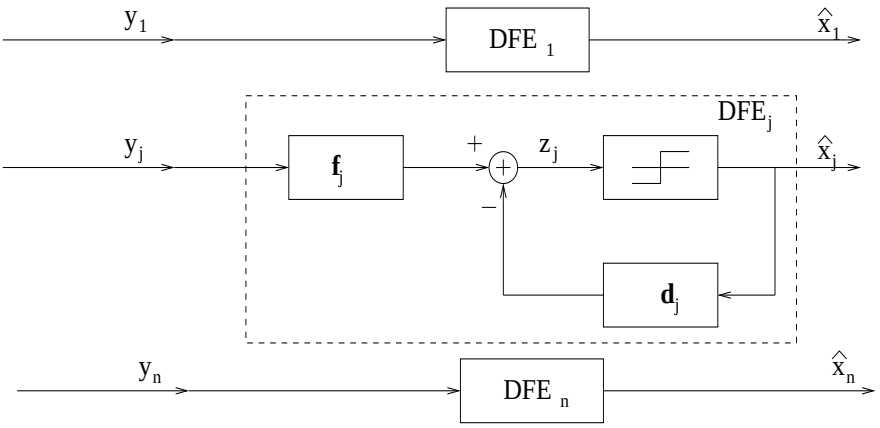

Figure 5.6: NC MIMO MMSE-DFE structure.

past N_d estimated symbols for each output we get the soft estimate:

$$\hat{z}_j(k) = \sum_{q=1}^{N_f} f_{jq} y_j(k-q) - \sum_{q=1}^{N_d} d_{jq} \hat{x}_j(k-q). \quad (5.37)$$

The FF and FB filters of the user j are obtained by minimizing the following cost functions with respect to $\mathbf{f}_j$ and $\mathbf{d}_j$:

$$\mathcal{J}_j = \mathrm{E}\{(x_j(k-\Delta) - \hat{z}_j(k))^2\}, \quad (5.38)$$

where δ is the equalization delay and $z_j(k)$ is given by:

$$z_j = \mathbf{x}_1^T \mathbf{H}_{1j} \mathbf{f}_j + \ldots + \mathbf{x}_i^T \mathbf{H}_{ij} \mathbf{f}_j + \ldots + \mathbf{x}_n^T \mathbf{H}_{nj} \mathbf{f}_j + \mathbf{v}_j^T \mathbf{f}_j - \hat{\mathbf{x}}_j^T \mathbf{d}_j. \tag{5.39}$$

In equation (5.38) we assume that past decisions are correct. Moreover, the input sequences are assumed to be uncorrelated with each other and with the noise.

Let us consider the observation vector received at receiver j. When we calculate the pair $(\mathbf{f}_j, \mathbf{d}_j)$, index j is fixed and i is running from $1, \ldots, n$. We use the notation $\hat{\mathbf{H}}_{ij} \mid_{i=j}$ for the case when the indexes i and j are equal. Hence, $\hat{\mathbf{H}}_{ii} = \hat{\mathbf{H}}_{ij} \mid_{i=j}$ corresponds to the direct path channel. Under the above assumptions and notation, for an $n \times m$ MIMO system we obtain:

$$\begin{cases} \mathbf{f}_j &= \left[\hat{\mathbf{H}}_{1j}^T \hat{\mathbf{H}}_{1j} + \ldots + \hat{\mathbf{H}}_{ij}^T \mid_{i=j} \mathbf{P}_{DFE}^j \hat{\mathbf{H}}_{ij} \mid_{i=j} + \ldots + \hat{\mathbf{H}}_{nj}^T \hat{\mathbf{H}}_{nj} + \lambda \mathbf{I}\right]^{-1} \hat{\mathbf{H}}_{ij}^T \mid_{i=j} \mathbf{e}_\Delta \\ \mathbf{d}_j &= \mathbf{M}_j \hat{\mathbf{H}}_{ij} \mid_{i=j} \mathbf{f}_j, \end{cases} \tag{5.40}$$

where $\mathbf{M}_j = (\mathbf{0}_{Nd \times \Delta} \;\; \mathbf{I}_{N_d \times N_d} \;\; \mathbf{0}_{N_d \times N_{ch} - N_d - \Delta})$, and $\mathbf{P}_{DFE}^j = (\mathbf{I} - \mathbf{M}_j^T \mathbf{M}_j)$. Furthermore $\lambda = \frac{1}{\sigma_v^2}$ and $\mathbf{e}_\Delta = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is the standard basis vector, with a value one at positions Δ, $0 \leq \Delta \leq N_f$. The derivation of the above two equations is presented in Paper V for a 2×2 case and in Paper VI for a general case.

Finally, the symbol estimate $\hat{x}_i$ at time k is obtained from:

$$\hat{x}_i(k) \;\; = \;\; \arg\min_{\alpha \in \mathcal{X}} |\alpha - \hat{z}_i(k)|, \tag{5.41}$$

where $\mathcal{X}$ is a finite alphabet.

5.5 Discussion

Using channel models based on measurements is very important when trying to simulate real communications environments. COST 207 is an effective model based on measurements of actual outdoor channels. These measurement-based power delay profiles and Doppler spectra of outdoor channels are useful in testing estimation and equalization algorithms. Different models have to be used if one wants to perform simulations in indoor environments, for example [93]. Nevertheless, for more advanced mathematical channel models one should consider recent COST [95] projects, such as COST 231, COST 259, COST 273 [96, 98, 101], as well as recent models proposed in technical journals, such as those in [3, 41, 135].

Joint channel tracking and equalization algorithms can perform well in the face of time-varying channels. In fact they outperform conventional adaptive equalization algorithms, such as LMS or RLS algorithms, which do not explicitly incorporate quasi-invariant channel statistics such as Doppler rate and the channel mean. This quantities may be known to the receiver as a result of prior processing. Channels which change rapidly can be tracked continuously. Hence, Kalman filters can be used as optimal channel estimators [14, 123]. The performance of a DFE coupled with LMS or RLS channel estimators was analyzed for a SISO channel model in [14, 117].

As shown in [128], fitting a crude model of channel time variation is better than no model at all. However, the complexity of the receiver increases relative to that of ordinary Kalman filtering techniques because multiple matrices of AR coefficients have to be estimated at every re-training of the algorithm. A good trade-off between complexity and performance is obtained using a simpler model. A low order AR model or even a simple Markov model can capture most of the channel tap dynamics and lead to effective tracking algorithms [83]. When using a Kalman filter for channel estimation and tracking, the system must include a noise estimation stage because noise statistics are not known *a priori* and may vary in time. Errors in channel estimation result in additional DFE performance degradation. This problem has been studied in [123].

Linear equalizers are best suited for channels which vary slowly. Deep frequency-selective fades are a characteristic of some common wireless channels. For such channels, DFEs are generally preferred to linear equalizers, since their complexity is comparable and their performance suffers less under amplitude distortion. Assuming perfect knowledge of the MIMO channel, the optimum receiver is a maximum likelihood sequence estimator (MLSE), but its complexity is prohibitive, even for low-order channels with a small number of inputs and outputs [124].

Equalization performed with a DFE gives a good trade-off between performance and complexity, compared to equalization with an optimal Viterbi algorithm. We believe that is an open discussion on advantages/disadvantages issues with respect to to fully-connected and non-connected MIMO MMSE-DFE structures. The equalizers are derived using different approaches, so it is difficult to make a fair comparison. However, the performance may depend on angular spread [135]. This implies that for mobile stations and for indoor base stations the angular spread

is high (almost $360°$) [135], thus low correlation can be achieved and NC MIMO DFE may give good results. In scenarios where the angular spread is low, and the spacing between antennas is not of a few wavelength in order to assure low correlation, a fully-connected MIMO DFE may perform better since it exploits this correlations. One drawback of the DFE structure may be the fact that errors may occur in bursts due to the fed-back symbol estimates and may lead to poor equalization.

Adaptive algorithms need acquisition and tracking phases. During acquisition, the initial values of the parameters are estimated. Usually acquisition is done with a known training sequence. Tracking may be needed in a time-varying channel to update the estimates obtained during the acquisition phase. The results from the literature in both SISO [14, 123] and MIMO [83] scenarios and also our simulation results validate the assumption that in the case of joint tracking and equalization algorithms, Kalman filter techniques offer the best channel tracking performance and DFE structure is a good tradeoff between complexity and performance.

Chapter 6

Summary

Blind source separation methods allow solution of many difficult signal processing problems in different application domains. Blind techniques are very important in applications for which there are sets of recorded data or observations, but little or no other information about specific parameter values of the system which produced the observations. An example is the measurement of multiple electroencephalogram traces, where sensor recordings are available, but no model for the brain signals which contribute to the traces. Blind techniques can produce very practical results. For instance, in communications, spectrum is a scarce resource and there is a need for higher data rates. Blind receivers require no training and allow transmission of user data in place of training sequences. On the other hand, communications receivers have to achieve high performance in demanding environments. Fully blind receivers may suffer from slow convergence and may have ambiguities. Hence, *semi-blind* algorithms have been introduced [35]. Semi-blind algorithms combine blind techniques and short training sequences, as well as other known information. Such algorithms can show improved performance relative to blind methods in challenging environments. They represent a practical alternative to training-based methods.

In this thesis a blind recursive method for solving the separation problem of linear instantaneous mixing was proposed. The possibilities of time-varying mixing matrices and noise were taken into account. The changes that may occur in the signal subspace were also considered and it was shown that they can be tracked on-line. A sample covariance matrix computed over a small window compared with the covariance matrix of the subspace revealed changes in the signal sub-

space. This was due to the high fluctuations in the correlation difference between the window estimate and the subspace covariance matrix. A simple thresholding was used to determine a change in environment and the subspace tracker was reinitialized. In a blind equalization application, fractional sampling was used to convert a finite input response multiple-input multiple-output (FIR-MIMO) model into a instantaneous multiple-input multiple-output (I-MIMO) model. The thesis showed that this technique allows recursive blind source separation to be applied to equalization of slowly time-varying channels.

The blind equalization work reported in this thesis has provided new solutions for the more challenging problem of MIMO equalization of time-varying channels. State-space models were useful for modeling time-varying systems. Recursive estimation was a natural way to deal with such models and was a key element of the methods developed in this work. Interestingly, the recursive techniques developed here were related to Kalman filter, which even four decades after its invention proves to be a valuable tool. Kalman filtering was the subject of active research in control theory in 60's. Demanding wireless communication applications of the $21st$ century may bring it back into the spotlight. There are still important aspects of Kalman filters which deserve closer attention. For example, solutions to the noise estimation problem can be very useful if one wants to have close to optimal algorithms which require little known information. In this work a solution for real-time noise estimation was presented based on testing the whiteness of innovations.

This thesis showed that a semi-blind technique can be used for equalization of MIMO time-varying channels. The algorithm derived for this purpose was based on a state-space model and recursive estimation. It worked in two stages, first estimating the channel and then equalizing it. A short training sequence was supplied for initializing a channel tracker. Subsequently, the algorithm worked in a decision-directed mode, meaning that the symbols from the equalization part were fed back and used to update the channel estimate. The channel estimation stage was based on Kalman filtering and included a noise estimation stage. This resulted in near-optimal channel estimation. The equalization was based on a non-connected MIMO minimum mean square error - decision feedback equalizer (MMSE-DFE) structure. In this structure, an independent DFE operated on each sensor. Simulation results showed that the algorithm was able to track changes

in time-varying channels and also to reduce intersymbol and interuser interference. It was also shown that the state and observation noise covariances can be successfully identified.

One possible topic of future research is closer analysis of the FIR-MIMO equalization problem. The channel tracking stage should be made robust to error bursts. These investigations may provide guidelines for the design of optimal training sequences that enable fast, accurate recovery following deep fades. The noise estimation stage needs careful understanding as it is critical to the performance of the Kalman filter. Erroneous estimates of the noise covariances can seriously impair Kalman filter behavior. Another important aspect of Kalman filtering is choice of the state transition matrix. Usually this matrix is assumed to be known and almost an identity matrix. Sometimes it is treated as having autoregressive parameters. However, more investigation is needed to determine if these assumptions are valid. A closer investigation of the interuser interference cancellation problem should be performed. The influence of user signal power on the performance of the non-connected MIMO MMSE-DFE should be investigated. Further solutions should be considered to improve the performance of the equalizer. For example, combining space-time coding with semi-blind equalization may lead to robust algorithms. Extension of blind methods applied to multicarrier systems would also be useful.

Appendix A

COST 207 model

The following four types of Doppler spectra are defined in the COST 207 model [100] and can be also found in [93]:

1. *CLASS* - classical Jakes Doppler spectrum used for paths with delays not in excess of 500 ns.

2. *GAUS 1* - is a sum of two Gaussian functions and is used for excess delay times in the range of 500 ns to 2 μs.

3. *GAUS 2* - is a sum of two Gaussian functions and is used for paths with delays equal or more than 2 μs.

4. *RICE* - is the sum of a classical Doppler spectrum and one direct path, such that the total multipath contribution is equal to that of the direct path. This spectrum is used for the shortest path of the model for propagation in RA.

The way in which these Doppler spectra should be applied to the four propagation classes is shown in Tables A.1-A.4.

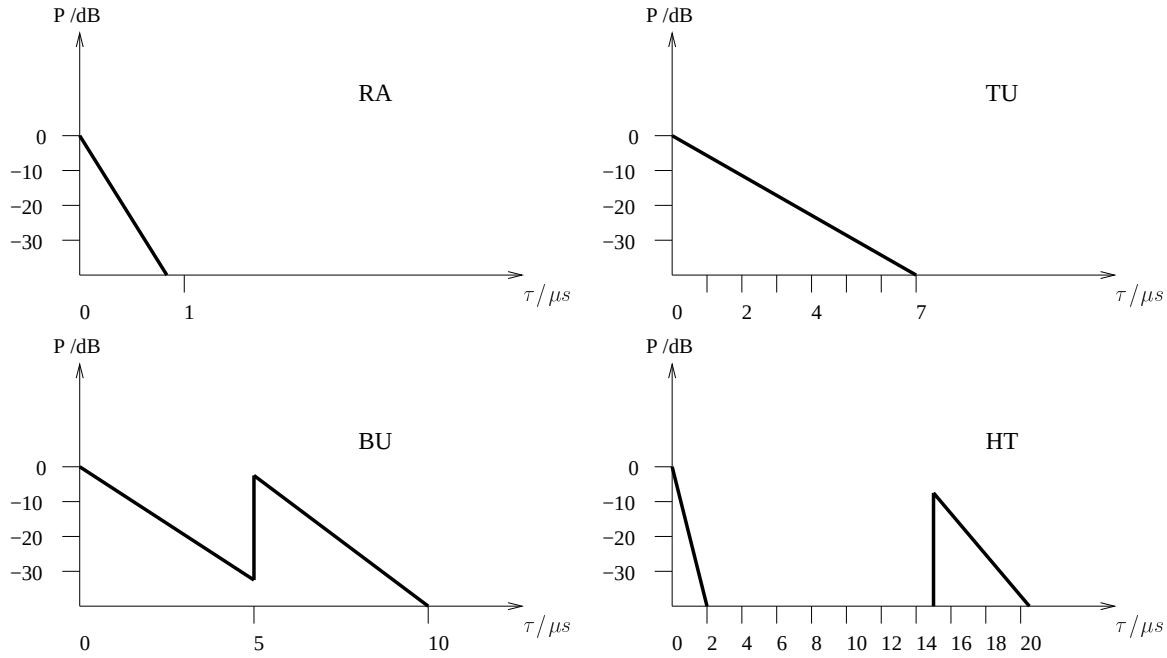

Figure A.1: COST-207 power delay profiles.

Tap #	Delay [μs]	Power [dB]	Doppler spectra
1	0	0	RICE
2	0.2	-2	CLASS
3	0.4	-10	CLASS
4	0.6	-20	CLASS

Table A.1: Parameters for Rural Area (RA) channel.

Tap #	Delay [μs]	Power [dB]	Doppler spectra
1	0	-3	CLASS
2	0.2	0	CLASS
3	0.6	-2	GAUS 1
4	1.6	-6	GAUS 1
5	2.4	-8	GAUS 2
6	5.0	-10	GAUS 2

Table A.2: Parameters for Typical Urban (TU) channel.

Tap #	Delay [μs]	Power [dB]	Doppler spectra
1	0	-3	CLASS
2	0.4	0	CLASS
3	1.0	-3	GAUS 1
4	1.6	-5	GAUS 1
5	5.0	-2	GAUS 2
6	6.6	-4	GAUS 2

Table A.3: Parameters for Bad Urban (BU) channel.

Tap #	Delay [μs]	Power [dB]	Doppler spectra
1	0	0	CLASS
2	0.2	-2	CLASS
3	0.4	-4	CLASS
4	0.6	-7	CLASS
5	15	-6	GAUS 2
6	17.2	-1	GAUS 2

Table A.4: Parameters for Hilly Terrain (HT) channel.

Bibliography

[1] Blind system identification and estimation. volume 86, pages 1901–2116. Proceedings of the IEEE, 1998.

[2] Independence and artificial neural networks. volume 22. Neurocomputing, 1998.

[3] Special issue on channel and propagation models for wireless system design I. volume 20, pages 493–647. IEEE Journal on Selected Areas in Communications, 2002.

[4] Special issue on Monte Carlo methods for statistical signal processing. volume 50, pages 173–449. IEEE Transactions on Signal Processing, 2002.

[5] N. Al-Dhahir and J.M. Cioffi. MMSE decision feedback equalizers: Finite length results. *IEEE Transactions on Information Theory*, 41:961–976, 1995.

[6] N. Al-Dhahir and A.H. Sayed. The finite length Multi-Input Multi-Output MMSE-DFE. *IEEE Transactions on Signal Processing*, 48(10):961–976, 2000.

[7] S-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

[8] S-I. Amari, T.-P. Chen, and A. Cichocki. Stability analysis of adaptive blind source separation. *Neural Networks*, 10(8):1345–1351, 1997.

[9] S-I. Amari and A. Cichocki. Adaptive blind signal processing-neural network approaches. *Proceedings of the IEEE*, 86(10):2026–2048, 1998.

[10] S-I. Amari, A. Cichocki, and H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Processing Systems*, 8:752–763, 1996.

[11] S-I. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, pages 752–763, 1995.

[12] S-I. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang. Multichannel blind deconvolution and equalization using the natural gradient. *Proceedings of the IEEE Workshop on Signal Processing and Advances in Wireless Communications*, pages 101–104, 1997.

[13] Sage A.P. and Husa G.W. Adaptive filtering with unknown prior statistics. *Proceedings of Joint Automatic Control Conference*, pages 760–769, 1969.

[14] Y. Ara and H. Ogiwara. Adaptive equalization of selective fading channel based on AR-model of channel-impulse-response fluctuation. *Electronics and Communications in Japan*, 74(7):76–86, 1991.

[15] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[16] H. Attias. Independent factor analysis. *Neural Computation*, 11:803–851, 1999.

[17] A. J. Bell and T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[18] P.A. Bello. Characterization of randomly time-variant linear channels. *IEEE Transactions on Communications Systems*, 11(4):360–393, 1963.

[19] D. Boss, K-D. Kammeyer, and T. Petermann. Is blind channel estimation feasible in mobile communications systems? a study based on GSM. *IEEE Journal on Selected Areas in Communications*, 16(8):1479–1492, 1998.

[20] J-F. Cardoso. Source separation using higher order moments. *Proceedings of ICASSP*, pages 2109–2112, 1989.

[21] J-F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, 1998.

[22] J-F. Cardoso. High-order contrasts for Independent Component Analysis. *Neural Computation*, 11(1):157–192, 1999.

[23] J-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44:3017–3030, 1996.

[24] J-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proceedings F*, 140(6):362–370, 1993.

[25] R.A. Casas, T.J. Endres, A. Touzni, C.R. Johnson Jr, and J.R. Treichler. Current approaches to blind decision feedback equalization. In G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors, *Signal Processing Advances in Wireless and Mobile Communications, Volume 1: Trends in Channel Estimation and Equalization*. Prentice Hall, 2000.

[26] R.A. Casas, P.B. Schniter, J. Balakrishnan, C.R. Johnson, Jr. Berg, and C.U. Berg. *DFE Tutorial*, 1998.

[27] C.K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications*. Springer, 3rd edition, 1999.

[28] A. Cichocki and S-I. Amari. *Adaptive Blind Signal and Image Processing : Learning Algorithms and Applications*. Wiley, 2002.

[29] A. Cichocki and L Zhang. Two-stage blind deconvolution using state-space models. *Proceedings of ICONIP*, pages 729–732, 1998.

[30] P. Comon. Independent Component Analysis – a new concept? *Signal Processing*, 36(3):287–314, 1994.

[31] P. Comon and P. Chevalier. *Unsupervised Adaptive Filtering, Vol. I: Blind Source Separation*, chapter Blind Source Separation: Models, Concepts, Algorithms, and Performance. John Wiley & Sons, 2000.

[32] P. Comon and G. H. Golub. Tracking a few extreme singular values and vectors in signal processing. *Proceedings of the IEEE*, 78(8):1327–1343, 1990.

[33] R. Cristescu, T. Ristaniemi, J. Joutsensalo, and J. Karhunen. Blind separation of convolved mixtures for CDMA systems. *European Signal Processing Conference, EUSIPCO*, pages 619–622, 2000.

[34] L.M. Davis, I.B. Collings, and R.J. Evans. Coupled estimators for equalization of fast-fading mobile channels. *IEEE Transactions on Communications*, 46(10):1262–1265, 1998.

[35] E. de Carvalho and D. Slock. Semi-blind methods for FIR multichannel estimation. In G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors, *Signal Processing Advances in Wireless and Mobile Communications, Volume 1: Trends in Channel Estimation and Equalization*. Prentice Hall, 2000.

[36] S. C. Douglas. Combined subspace tracking, prewhitening, and contrast optimization for noisy blind signal separation. *Proceedings of the Independent Component Analysis and Blind Signal Separation Workshop (ICA 2000)*, pages 579–584, 2000.

[37] S. C. Douglas. Numerically-robust adaptive subspace tracking using householder transformations. *IEEE Sensor Array Multichannel Workshop*, pages 499–503, 2000.

[38] S. C. Douglas. *Handbook of Neural Network Signal Processing*, chapter Blind Signal Separation and Blind Deconvolution. CRC Press, 2001.

[39] S. C. Douglas, A. Cichocki, and S-I. Amari. Multichannel blind separation and deconvolution of sources with arbitrary distributions. *IEEE Workshop on Neural Networks for Signal Processing*, pages 436–445, 1997.

[40] S.C. Douglas and S. Haykin. *Unsupervised Adaptive Filtering, Vol. II: Blind Deconvolution*, chapter Relationships between Blind Deconvolution and Blind Source Separation. John Wiley & Sons, 2000.

[41] R.B. Erdel, P. Cardieri, K.W. Sowerby, and T.S. Rappaport. Overview of spatial channel models for antenna array communications. *IEEE Personal Communications Magazine*, 5(1):10–22, 1998.

[42] R. Everson and J. S. Roberts. Non-stationary Independent Component Analysis. *International Conference on Artificial Neural Networks*, pages 503–508, 1999.

[43] R. Everson and J. S. Roberts, editors. *Advances in Independent Components Analysis*. Kluwer Academic Publishers, 2000.

[44] R. Everson and J. S. Roberts. *Independent Component Analysis Principles and Practice*. Cambridge University Press, 2001.

[45] S.A. Fechtel and M. Meyr. An investigation of channel estimation and equalization techniques for moderately rapid fading HF-channels. *Proceedings of the International Conference on Communications*, pages 768–772, 1991.

[46] A. B. Gharbi and F. M. Salam. Algorithms for blind signal separation and recovery in static and dynamic environments. *IEEE International Symposium on Circuits and Systems*, 1:713–716, 1997.

[47] G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors. *Signal Processing Advances in Wireless and Mobile Communications, Volume 1: Trends in Channel Estimation and Equalization.* Prentice Hall, 2000.

[48] X. Giannakopoulos, J. Karhunen, and E. Oja. An experimental comparison of neural ICA algorithms with real-world data. *Proceedings of the International Joint Conference on Neural Networks (IJCNN'99)*, pages 888–893, 1999.

[49] M. Girolami. An alternative perspective on adaptive Independent Component Analysis algorithms. *Neural Computation*, 10(8), 1998.

[50] M. Girolami. Noise reduction and speech enhancement via temporal anti-hebbian learning. *Proceedings of ICASSP*, 1998.

[51] M. Girolami, editor. *Self-Organising Neural Networks - Indepedent Component Analysis and Blind Source Separation*. Springer-Verlag, 1999.

[52] D.N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Transactions on Communications*, 28(11):1867–1875, 1980.

[53] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, pages 107–113, 1993.

[54] A. Gorokhov, P. Loubaton, and E. Moulines. Second order blind equalization in multiple input multiple output FIR systems: a weighted least squares approach. *Proceedings of ICASSP*, 5:2415–2418, 1996.

[55] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, 3rd edition, 1996.

[56] S. Haykin, editor. *Unsupervised Adaptive Filtering, Vol. I, Blind Source Separation*. Wiley, 2000.

[57] S. Haykin, editor. *Unsupervised Adaptive Filtering, Vol. II, Blind Deconvolution*. Wiley, 2000.

[58] J. Herault and C. Jutten. Blind separation of sources, part II: Problem statement. *Signal Processing*, 24(1):11–20, 1991.

[59] J. Herault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. *Proc. Xéme colloque GRETSI*, pages 1017–1022, 1985.

[60] P. Hoeher. A statistical discrete-time model for the WSSUS multipath channel. *IEEE Transactions on Vehicular Technology*, 41(4):461–468, 1992.

[61] F.M. Hsu. Square root Kalman filtering for high-speed data received over fading dispersive HF channels. *IEEE Transactions on Information Theory*, 28(5):753–763, 1982.

[62] Y. Hua. Fast maximum likelohood for blind identification of multiple FIR channels. *IEEE Transactions on Signal Processing*, 44(3):661–672, 1996.

[63] A. Hyvärinen. A family of fixed-point algorithms for Independent Component Analysis. *Proceedings of ICASSP*, pages 3917–3920, 1997.

[64] A. Hyvärinen. Fast and robust fixed-point algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[65] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.

[66] R.A. Iltis. Joint estimation of PN code delay and multipath using the extended Kalman filter. *IEEE Transactions on Communications*, 38(10):1677–1685, 1990.

[67] W.C. Jakes. *Microwave Mobile Communications*. Wiley, 1974.

[68] C. Jutten and J. Herault. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[69] Myers K.A. and B.D. Tapley. Adaptive sequential estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, 4(21):520–523, 1976.

[70] I. Karasalo. Estimating the covariance matrix by signal subspace averaging. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(1):8–12, 1986.

[71] J. Karhunen, A. Cichocki, W. Kasprzak, and P. Pajunen. On neural blind separation with noise suppression and redundancy reduction. *International Journal of Neural Systems*, 8(2):219–237, 1997.

[72] J. Karhunen and J. Joutsensalo. Representation and separation of signals using nonlinear PCA type learning. *Neural Networks*, 1(8):113–127, 1994.

[73] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo. A class of neural networks for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 8(3):486–504, 1997.

[74] J. Karhunen and P. Pajunen. Blind source separation and tracking using nonlinear PCA criterion: A least-squares approach. *IEEE International Conference on Neural Network*, pages 2147–2152, 1997.

[75] J. Karhunen and P. Pajunen. Blind source separation using least-squares type adaptive algorithms. *Proceedings of the ICASSP*, pages 3361–3364, 1997.

[76] J. Karhunen, P. Pajunen, and E. Oja. The nonlinear PCA criterion in blind source separation: relations with other approaches. *Neurocomputing*, 22:5–21, 1997.

[77] B. Karlsen, J. Larsen, B.D.H. Sørensen, and B.K. Jakobsen. Comparison of PCA and ICA based clutter reduction in GPR systems for anti-personal landmine detection. *IEEE Workshop on Statistical Signal Processing*, pages 146–149, 2001.

[78] J. Karvanen, J. Eriksson, and V. Koivunen. Adaptive score functions for maximum likelihood ICA. *Journal of VLSI Signal Processing Systems*, 2002. To appear.

[79] V. Koivunen, J. Laurila, and E. Bonek. Blind methods for wireless communication receivers. *Review of Radio Science 1999-2002*, 2002. To appear.

[80] V. Koivunen and E. Oja. Predictor-corrector structure for real-time blind separation from noisy mixtures. *Proceedings of the First International Conference on Independent Component Analysis and Blind Source Separation: ICA'99*, pages 479–484, 1999.

[81] C. Komninakis, C. Fragouli, A.H. Sayed, and R.D. Wesel. Channel estimation and equalization in fading. *Proceedings of 33rd Asilomar Conference on Signals, Systems, and Computers*, 2:1159 –1163, 1999.

[82] C. Komninakis, C. Fragouli, A.H. Sayed, and R.D. Wesel. Adaptive Multi-Input Multi-Output fading channel equalization using Kalman estimation. *IEEE International Conference on Communications*, 3:1655–1659, 2000.

[83] C. Komninakis, C. Fragouli, A.H. Sayed, and R.D. Wesel. Multi-Input Multi-Output fading channel tracking and equalization using Kalman estimation. *IEEE Transactions on Signal Processing*, 50(5):1065–1076, 2002.

[84] R.H. Lambert. *Multichannel Blind Deconvolution: FIR matrix algebra and separation of multipath mixtures*. PhD thesis, University of Southern California, 1996.

[85] J. Laurila. *Semi-blind Detection of Co-Channel Signals in Mobile Communications*. PhD thesis, Technische Universität Wien, 2000.

[86] T.-W. Lee. *Independent Component Analysis - Theory and Applications*. Kluwer, 1998.

[87] E.L. Lehmann. *Nonparametrics, Statistical Methods Based on Ranks*. Prentice-Hall, 1998.

[88] J.C. Liberti and T.S. Rappaport. *Smart Antennas for Wireless Communications*. Prentice Hall, 1999.

[89] P. Loubaton, E. Moulines, and P. Regalia. Subspace method for blind identification and deconvolution. In G.B. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors, *Signal Processing Advances in Wireless and Mobile Communications, Volume 1: Trends in Channel Estimation and Equalization*. Prentice Hall, 2000.

[90] H. Mathis and S.C. Douglas. On the existence of universal nonlinearities for blind source separation. *IEEE Transactions on Signal Processing*, 50(5):1007–184, 2002.

[91] P.S. Maybeck. *Stochastic models, estimation, and control*, volume 1,2,3. Academic Press, 1982.

[92] R.K. Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(2):175–184, 1970.

[93] A.F. Molisch, editor. *Wideband wireless digital communications*. Prentice Hall, 2000.

[94] E. Moreau and O. Macchi. New self-adaptive algorithms for source separation based on contrast functions. *Proceedings of IEEE Signal Processing Workshop on Higher Order Statistics*, pages 215–219, 1993.

[95] Commission of the European Communities. Telecommunication information science and technology. http://www.cordis.lu/cost.

[96] Commission of the European Communities. Evolution of land mobile radio (including personal communication). Office for Official Publications of the European Communities, Luxembourg, Apr. 1989 - Apr. 1996. http://www.lx.it.pt/cost231/.

[97] Commission of the European Communities. Integrated space/terrestrial mobile networks. Office for Official Publications of the European Communities, Luxembourg, Apr. 1991 - Apr. 1995. http://www.estec.esa.nl/xewww/cost227/cost227.html.

[98] Commission of the European Communities. Wireless flexible personalized communications. Office for Official Publications of the European Communities, Luxembourg, Dec. 1996 - Apr. 2000. http://www.lx.it.pt/cost259.

[99] Commission of the European Communities. Propagation impairment mitigation for millimetre wave radio systems. Office for Official Publications of the European Communities, Luxembourg, Jun. 2001 - Jun. 2005. http://www.cost280.rl.ac.uk/.

[100] Commission of the European Communities. Digital land mobile radio communications - COST 207. Office for Official Publications of the European Communities, Luxembourg, Mar. 14, 1984 - Sept. 13, 1988.

[101] Commission of the European Communities. Towards mobile broadband multimedia networks. Office for Official Publications of the European Communities, Luxembourg, May 2001 - May 2005. http://www.lx.it.pt/cost273.

[102] E. Oja. Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1:61–68, 1989.

[103] E. Oja. The nonlinear PCA learning rule and Independent Component Analysis - mathematical analysis. *Neurocomputing*, 17:25–45, 1997.

[104] E. Oja, H. Ogawa, and J. Wangviwattana. *Learning in nonlinear constrained Hebbian networks*. Elsevier, 1991.

[105] H.C. Papadopoulos. *Wireless Communications: Signal Processing Perspectives*, chapter Equalization of Multiuser Channels. Prentice Hall, 1998.

[106] A.J. Paulraj and C.B. Papadias. Space-time processing for wireless communications. *IEEE Signal Processing Magazine*, 14:49–83, 1997.

[107] D.-T. Pham and P. Garat. Blind separation of mixture of independent sources through a quasimaximum likelohood approach. *IEEE Transactions on Signal Processing*, 45:1712–1725, 1997.

[108] H.V. Poor and G.W. Wornell, editors. *Wireless Communications: Signal Processing Perspectives*. Prentice Hall, 1998.

[109] J.G. Proakis. *Digital Communications*. McGraw Hill, 3rd edition, 1999.

[110] T.S. Rappaport. *Wireless Communications: Principles Practice*. Prentice Hall, 1999.

[111] E.C. Real, D.W. Tufts, and J.W. Cooley. Two algorithms for fast approximate subspace tracking. *IEEE Transactions on Signal Processing*, 47(7):1936–1945, 1999.

[112] S. Roberts and R. Everson, editors. *Independent Component Analysis Principles and Practice*. Cambridge, 2001.

[113] S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11(2):305–345, 1999.

[114] F. M. Salam. An adaptive network for blind separation of independent signals. *IEEE International Symposium on Circuits and Systems*, 1:1431–1434, 1993.

[115] Y. Sato. A method of self recovering equalization for multilevel amplitude modulation. *IEEE Transactions on Communications*, 23:679–682, 1975.

[116] O. Shalvi and E. Weinstein. New criteria for blind deconvolution of nonminimum phase systems (channels). *IEEE Transactions on Information Theory*, pages 312–321, 1990.

[117] P. Shukla and L. Turner. Channel-estimation-based adaptive DFE for fading multipath radio channels. *IEE Proceedings I*, 138(6):525–543, 1991.

[118] M. Sirbu, M. Enescu, and V. Koivunen. Estimating noise statistics in adaptive semi-blind equalization. *Proceedings of 35th Asilomar Conference on Signals, Systems, and Computers*, 2001.

[119] M. Sirbu, M. Enescu, and V. Koivunen. Real-time semi-blind equalization of time varying channels. *IEEE-Eurasip Workshop on Nonlinear Signal and Image Processing*, pages 90–100, 2001.

[120] B. Sklar. Rayleigh fading channels in mobile digital communication systems Part I: characterization. *IEEE Communications Magazine*, pages 90–100, 1997.

[121] J.E. Smee and S.C. Schwartz. Adaptive feedforward/feedback architectures for multiuser detection in high data rate wireless CDMA networks. *IEEE Transactions on Communications*, 48(6):996–1011, 2000.

[122] M.R. Spiegel, editor. *Theory and Problems of Statistics*. McGraw-Hill, 2 edition, 1994.

[123] M. Stojanovic, J.G. Proakis, and J.A. Catipovic. Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels. *IEEE Transactions on Communications*, 43(2/3/4):877–886, 1995.

[124] D.P. Taylor, G.M. Vitetta, B.D. Hart, and A. Mämmelä. Wireless channel equalization. *European Transactions on Telecommunications*, 9(2):117–143, 1998.

[125] M. Toeltsch, J. Laurila, K. Kalliola, A.F. Molisch, P. Vainikainen, and E. Bonek. Statistical characterization of urban spatial radio channels. *IEEE Journal on Selected Areas in Communications*, 20(3):539–549, 2002.

[126] L. Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelohood methods. *Proceedings of the IEEE*, 86(10):1951–1968, 1998.

[127] K. Torkkola. Blind separation for audio signals - are we there yet? *Proceedings of the First International Conference on Independent Component Analysis and Blind Source Separation: ICA'99*, pages 239–244, 1999.

[128] M.K. Tsatsanis, G.B. Giannakis, and G. Zhou. Estimation and equalization of fading channels with random coefficients. *Signal Processing*, 53:211–229, 1996.

[129] M.K. Tsatsanis, G.B. Giannakis, and G. Zhou. Estimation and equalization of fading channels with random coefficients. *Proceedings of ICASSP*, 53:1093–1096, 1996.

[130] D.W. Tufts, E.C. Real, and J.W. Cooley. Fast approximate subspace tracking (FAST). *Proceedings of ICASSP*, 1:547–560, 1997.

[131] J.K. Tugnait. Blind equalization and channel estimation for multiple-input multiple-output communications systems. *Proceedings of ICASSP*, 5:2443–2446, 1996.

[132] J.K. Tugnait, L. Tong, and Z. Ding. Single-user channel estimation and equalization. *IEEE Signal Processing Magazine*, 17:17–28, 2000.

[133] R. Vigario. Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph. clin. Neurophysiol.*, 3(103):395–404, 1997.

[134] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.

[135] J.H. Winters. Smart antennas for wireless systems. *IEEE Personal Communications*, 5(1):23–27, 1998.

[136] B Yang. Projection Approximation Subspace Tracking. *IEEE Transactions on Signal Processing*, 43(1):95–107, 1995.

[137] H.H. Yang. On-line blind equalization via on-line blind separation. *Signal Processing*, 68:271–281, 1998.

[138] H.H. Yang and S. Amari. Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information. *Neural Computation*, 9:1457–1482, 1997.

[139] J. Yang and S. Roy. On joint transmitter and receiver optimization for Multiple-Input-Multiple-Onput (MIMO) transmission systems. *Neural Computation*, 42(12):3221–3231, 1994.

[140] L. Zhang and A. Cichocki. Blind deconvolution / equalization using state-space models. *Proceedings of the IEEE workshop on NNSP*, pages 123–131, 1998.

[141] L. Zhang and A. Cichocki. Blind separation of filtered sources using state-space approach. *Advances in Neural Information Processing Systems*, 11:648–654, 1999.

[142] L. Zhang and A. Cichocki. Blind deconvolution of dynamical systems: A state-space approach. *Signal Processing*, 4(2):110–130, 2000.

[143] Y. Zhang and S.A. Kassam. Blind separation and equalization using fractional sampling of digital communications signals. *Signal Processing*, 81(12):2591–2608, 2001.

[144] Y. Zhang, J. Mannerkoski, V. Koivunen, and S.A. Kassam. Blind equalization for GSM systems using blind source separation. *submitted to IEEE Transactions on Communications*, 2001.

Publications