CLOSED-FORM HIGHER-ORDER ESTIMATORS FOR BLIND SEPARATION OF INDEPENDENT SOURCE SIGNALS IN INSTANTANEOUS LINEAR MIXTURES

THESIS SUBMITTED

IN ACCORDANCE WITH THE REQUIREMENTS OF

THE UNIVERSITY OF LIVERPOOL

FOR THE DEGREE OF

DOCTOR IN PHILOSOPHY

By

Vicente Zarzoso Gascón-Pelegrí

Department of Electrical Engineering and Electronics

The University of Liverpool

Liverpool, United Kingdom

October 1999

Declaration

I declare that this thesis embodies my own research work, and that it is composed by myself. Where appropriate, I have made acknowledgement to the work of others.

Vicente Zarzoso Gascón-Pelegrí

A mis padres,

Juan José y María Teresa,

a mis abuelos

Teresa, Vicente y Victoria,

y en memoria de mi abuelo

Matías.

Abstract

Blind source separation (BSS) aims to recover a set of unobservable independent source signals from the observation of a set of sensor outputs which contains mixtures of the sources. This thesis focuses on the closed-form estimators for the higher-order stage of the BSS of two sources from two noiseless instantaneous linear mixtures. Emphasis is laid on the notion that different estimators may offer dissimilar finite-sample performance, so that the statistical characterization of these methods is one of the main objectives of this work.

The approximate maximum likelihood (AML) estimator by Harroy and Lacoume is generalized to virtually any source probability distribution. The new extended expression — *extended ML (EML)* — is based on a scatter-plot complex centroid which depends on the fourth-order statistics of the data, and which allows a geometric interpretation inspired by Bogner and Clarke. The EML performance is studied and an analytic expression for the estimator's asymptotic probability density function (pdf) is determined. With the analysis tools we develop, another closed-form formula by Comon (CF) is understood and linked to the AML and the EML. The latter is the best option.

The EML estimator is the closed-form solution to certain optimization problems, which allows the extension of a contrast function derived by Moreau, and the simplification of the associated analytic solution derived by Comon-Moreau.

AML, EML and CF suffer a performance degradation when a source statistic approaches zero. To overcome this shortcoming, another fourth-order estimator — *alternative EML (AEML)* — is proposed, and a combined estimation strategy is adopted. A suboptimal rule to decide between the EML and the AEML is derived and validated.

Extensions to scenarios of more than two signals are implemented, based on a Jacobi-like iteration strategy originally proposed by Comon. The resulting methods are successfully applied to a real problem: the extraction of the fetal electrocardiogram from maternal cutaneous potential recordings.

The compact centroid-based expressions permit the formulation of straightforward adaptive versions, whose high convergence speed and global convergence under mild conditions are remarkable.

A closed-form estimation family based on the data nth-order statistics is unveiled, which yields the EML at $n = 4$. At $n = 3$, a novel third-order estimator — *third-order blind signal estimator (TOBSE)* — is obtained and analyzed.

Through the so-called *bicomplex numbers*, some of the previous results are extended to complex-valued mixtures, evidencing an elegant connection between the real and the complex case.

Numerous computer experiments verify the theoretical results, make comparisons among the techniques considered, and contrast them with other non-analytic procedures. The computational complexity of the methods is also discussed.

In conclusion, this investigation provides the first unified comprehensive vision of closed-form higher-order estimators for BSS.

Acknowledgements

First of all, I want to thank my supervisor, Prof. Asoke K. Nandi, for the independence he has given me throughout all this project, while always being on hand whenever I needed his help, guidance and enthusiasm.

Most of this research was carried out at the Signal Processing Division (SPD), Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow. Working there was a truly enjoyable experience, for which I am grateful to all the *BPD* members whom I had the chance to meet during my stay in Scotland.

In Glasgow, this work was mainly supported by a scholarship from the University of Strathclyde. Over the first year, support also came from the Defence Evaluation and Research Agency (DERA) of the UK, and later, from the Engineering and Physical Sciences Research Council of the UK.

The last part of this work was accomplished at the Department of Electrical Engineering and Electronics, The University of Liverpool. I want to express my gratitude to the many people who made our move to this Department more bearable. Kev Keller deserves a special "ooh la la!" mention, for his invaluable help at the beginning of our stay in Liverpool, for the close squash games, and for the mouth-watering barbecues. Derek, Giles and Lee have been great fellows too since we arrived in the English city.

The move from Glasgow to Liverpool would not have been so tolerable without the company of three other mates: Lindsay B. Jack, Frank Herrmann and June-Yule Lee. The former two, in addition, had to put up with my stress all over the final part of my PhD studies. I am aware of how enduring this experience can be, and so special thanks go to them too. Also, Lindsay Jack's help with the computer system was certainly invaluable.

It was hard to leave behind so many good friends in Glasgow, including my last flatmates, and best friends, Juan Antonio Pons Puig and Oscar Lázaro de Barrio. Their heart is even greater than the fun we have had together and, ever since we met, they have been showing me what friendship is all about. My lively discussions with Raquel Morera Sempere are greatly missed as well. A super sweet "xie xie" goes to Yin Yin Lam, for her special friendship and her support over the final stage of my PhD.

Thanks are also due to my brother Juanjo. His strong faith, reassuring trust and constant encouragement have been priceless, especially when facing difficult decisions.

On the technical side, I must start by sending my gratitude to Ira J. Clarke, from DERA-Malvern, whose work and ideas have inspired much of this investigation. I also want to mention Pierre Comon and Jean-François Cardoso, in my opinion the two greatest researchers on blind source separation, for their legendary contributions to this exciting problem. They have been a constant source of inspiration to me. I am thankful for the stimulating, insightful discussions I had the opportunity to enjoy with them, which led to many of the results presented in this thesis. I am also indebted to Øyvind Strømme, who kindly offered himself to read the original manuscript, appraised it very positively, and provided some interesting constructive suggestions.

Finally, I want to thank very specially all my family. I am infinitely grateful for the values they have passed down to me, and for their continuous support and encouragement throughout all my studies, which successfully culminate in this thesis.

Vicente Zarzoso Gascón-Pelegrí

Liverpool, October 1999.

Contents

List of Figures

List of Tables

Mathematical Notations

Blind separation system:

$\widehat{B}^{\dagger}$	whitening matrix
G, g_{ij}	global matrix, global-matrix element: $G \triangleq WM$
h_{ij}	linear time-invariant system
$H_{ij}\{\cdot\}$	transformation
M, $\mathbf{m}_i$, m_{ij}	mixing matrix, mixing-matrix column vector (transfer vector, source direction or source signature), mixing-matrix element
$\widehat{M}$, $\widehat{\mathbb{M}}$	mixing-matrix estimate, mixing-matrix estimator
$\mathbf{n}$, n_i, N	noise vector, ith noise signal, noise matrix
p	number of observations
q	number of sources
Q	unitary transformation
$\widehat{Q}$, $\widehat{\mathbb{Q}}$	unitary-transformation estimate, unitary-transformation estimator
$\mathbf{s}$, s_i, S	separator-output vector, ith separator output, separator output matrix
T	number of samples (sample size)
W, $\mathbf{w}_i$, w_{ij}	separating matrix, separating vector, separating-matrix element
$\mathbf{x}$, x_i, X	source vector, ith source signal, source-signal matrix
$\hat{\mathbf{x}}$, $\hat{x}_i$, $\hat{X}$	estimated source vector, ith estimated source signal, estimated source matrix
$\tilde{\mathbf{x}}$	normalized source vector
$\mathbf{y}$, y_i, Y	observed-mixture vector, ith observed signal, sensor-output matrix
$\mathbf{z}$, z_i, Z	whitened vector, ith whitened signal, whitened-output matrix

(*Mathematical Notations*, continued)

Statistics:

$\mathrm{E}[\cdot]$	mathematical expectation
$\mathrm{Cov}[\cdot]$	covariance
$\mathrm{Cum}[\cdot]$, κ	cumulant
$\mathrm{Mom}[\cdot]$, μ	moment
$\mathrm{N}(m, \sigma^2)$	Gaussian (normal) distribution of mean m and variance σ^2
p, $p(\cdot)$, P	probability, probability density function, power
R_x	spatial covariance matrix
$\mathrm{U}(a, b)$	uniform distribution between a and b
$\mathrm{Var}[\cdot]$	variance
γ, γ^+	source kurtosis sum, absolute source kurtosis sum
η, η^+	source kurtosis difference, absolute source kurtosis difference
κ, $\tilde{\kappa}$	cumulant, normalized cumulant
$\kappa_{(r)}$	rth-order cumulant tensor
μ, $\tilde{\mu}$	moment, normalized moment
σ^2, σ_N^2	variance, noise variance
ϕ, ψ	first, second characteristic function
$\hat{(\cdot)}$	estimate, sample estimate, approximation
$\xrightarrow{\mathrm{d}}$	convergence in distribution
$\xrightarrow{\mathrm{wp1}}$	convergence with probability one

Basic matrices and matrix operators:

D	invertible diagonal matrix
I_n	$n \times n$ identity matrix
P	permutation matrix
$(\cdot)^*$	complex-conjugate operator
$(\cdot)^{\mathrm{H}}$	hermitian (complex-conjugate transpose) operator
$(\cdot)^{-1}$	inverse

$(\cdot)^{\dagger}$	Moore-Penrose pseudoinverse operator
$(\cdot)^{\mathrm{T}}$	transposition operator

Rotation and unitary matrices:

Q	rotation or unitary matrix: $QQ^{\mathrm{H}} = Q^{\mathrm{H}}Q = I$
$\widehat{Q}$	estimated rotation
$\widehat{\mathbb{Q}}$	rotation-matrix estimator
$\mathcal{Q}(\theta)$	real Givens rotation matrix of angle θ: $\mathcal{Q}(\theta) = \mathcal{Q}(\theta,\, 0)$
$\mathcal{Q}(\theta,\, \alpha)$	complex Givens rotation matrix of angle θ and phase α

Scatter plots and centroids:

r	scatter-plot point modulus
ζ	transformed output scatter-plot point
$\nu,\, \nu'$	whitened-signal scatter-plot point, source scatter-plot point
$\xi,\, \xi'$	fourth-order centroid, alternative fourth-order centroid
ξ_{γ}	ξ times $\mathrm{sign}(\gamma)$
$\hat{\xi}_x$	source-centroid sample estimate
$\xi_n(\cdot)$	nth-order generalized centroid
$\xi'_n(\cdot)$	nth-order alternative generalized centroid
$\sigma_1^2,\, \sigma_2^2,\, c_{12}$	variance of ω_1, variance of ω_2, covariance of ω_1 and ω_2
$\tilde{\sigma}_1^2,\, \tilde{\sigma}_2^2,\, \tilde{c}_{12}$	variance of $\tilde{\omega}_1$, variance of $\tilde{\omega}_2$, covariance of $\tilde{\omega}_1$ and $\tilde{\omega}_2$
$\phi,\, \phi'$	whitened-signal, source scatter-plot point phase
$\omega_1,\, \omega_2$	real, imaginary part of source-centroid sample estimate $\hat{\xi}_x$
$\tilde{\omega}_1,\, \tilde{\omega}_2$	real, imaginary part of transformed source scatter-plot point

(*Mathematical Notations*, continued)

Angle and phase estimation:

$\mathrm{arctg}(\cdot)$	arc (inverse) tangent function
$\arg(\cdot)$, $\angle\cdot$	argument (phase) function
$\mathrm{tg}(\cdot)$	tangent function
α, $\hat{\alpha}$	true phase, estimated phase
δ	centroid phase
$\Delta\hat{\theta}$	angle-estimate error: $\Delta\hat{\theta} \triangleq \hat{\theta} - \theta$
θ, $\hat{\theta}$	true angle, estimated angle
Ω_φ	length-φ centered interval: $\Omega_\varphi \triangleq [-\varphi/2, \varphi/2]$

Complex and bicomplex numbers:

$\arg(\cdot)$, $\angle\cdot$	complex argument function
$\mathrm{barg}(\cdot)$	bicomplex argument function
$\mathrm{e}_\alpha^{\jmath\theta}$	bicomplex exponential: $\mathrm{e}_\alpha^{\jmath\theta} \triangleq \cos\theta + \jmath\mathrm{e}^{j\alpha}\sin\theta$
$\mathrm{I\!Im}(\cdot)$	imaginary part
$\mathbb{I}\mathrm{m}(\cdot)$	bicomplex imaginary (bimaginary) part
j	imaginary unit: $j^2 \triangleq -1$
$\jmath$	bicomplex imaginary (bimaginary) unit: $\jmath^2 \triangleq -1$
$\mathrm{I\!Re}(\cdot)$,	real part
$\mathbb{R}\mathrm{e}(\cdot)$,	bicomplex real (breal) part
$\bar{x}$	bicomplex number: $\bar{x} \triangleq a + \jmath b, \quad a, b \in \mathbb{C}$
$(\cdot)^*$	complex or bicomplex conjugate operator

Cost or contrast functions:

I, J, Ψ	cost or contrast function
J_{m}	cost function to be minimized
J_{M}	cost function to be maximized

(*Mathematical Notations*, continued)

Performance indices:

ISR	interference-to-signal ratio
SNR	signal-to-noise ratio
ϵ	mixing-matrix gap
ε^2	waveform similarity
ϱ	rejection ratio
Υ, $\tilde{\Upsilon}$	statistical independence, distance to independence

Multi-reference adaptive noise cancellation system:

a	primary input
n, N	number of reference inputs, number of filter taps
$P^{(i)}$, $\hat{P}^{(i)}$	ith-signal power, ith-signal estimated power
$r^{(i)}$, $\mathbf{r}^{(i)}$	ith reference signal, ith reference-signal vector
$w^{(i)}$, $\mathbf{w}^{(i)}$	ith-input filter coefficient, ith-input filter vector
ε	output signal
$\mu^{(i)}$	ith-input adaption coefficient
$\tilde{\mu}^{(i)}$	ith-input power adaption coefficient

Sets and groups:

$\mathcal{B}$	bicomplex numbers
$\mathcal{B}_u$	unitary bicomplex numbers
$\mathbb{C}$	complex numbers
$\mathrm{GL}(n, \mathbb{X})$	general linear group of $n \times n$ matrices over $\mathbb{X}$ (set of $n \times n$ matrices with entries in $\mathbb{X}$ and non-zero determinant)
$\mathbb{N}$	non-negative integer (whole) numbers: $\mathbb{N} \triangleq \{0, 1, 2, \dots\}$
$\mathbb{N}^+$	positive integer (natural) numbers: $\mathbb{N}^+ \triangleq \{1, 2, \dots\}$

(*Mathematical Notations, sets and groups*, continued)

$\mathcal{O}$	orthogonal matrices
$\mathcal{O}_u$	unitary matrices
$\mathbb{R}$	real numbers
$\mathrm{SL}(n, \mathbb{X})$	special linear group of $n \times n$ matrices over $\mathbb{X}$ (set of $n \times n$ matrices with entries in $\mathbb{X}$ and unit determinant)
$\mathbb{Z}$	integer numbers
$[\cdot]$	equivalence class of valid solutions

Relations and operators:

mod	modulo operator (remainder of integer division)
$\mathrm{O}(\cdot)$	order of magnitude
$*$	convolution operator
$\lfloor \cdot \rfloor$	floor operator (round off)
∇	gradient operator
∂	partial derivative
$\in$	membership
$\subset$	subset
$\approx$	approximately equal
$\underset{\substack{\uparrow \\ \cdots}}{=}$	equal because ...
$\triangleq$	equal by definition
$\equiv$	separation-solution equivalence
$\simeq$	isomorphism
$\propto$	proportionality
$\sim$	similarity
$\rightarrow$	convergence
$\mapsto$	mapping
$\Leftrightarrow$	if, and only if

(*Mathematical Notations*, continued)

Miscellaneous:

k	time index, iteration number
λ, μ	adaption coefficients (step sizes)

Acronyms

ACF	alternative CF
adAEML	adaptive AEML
adcombEML	adaptive combined EML
adEML	adaptive EML
AEML	alternative EML
AFOC	alternative fourth-order centroid
AML	approximate maximum likelihood
ARMA	auto-regressive moving average
AROT	adaptive rotation
ARS	absolute-rotation sum
BSS	blind source separation
CAEML	complex AEML
CCF	complex CF
CEML	complex EML
CF	Comon's formula
CLT	central limit theorem
combEML	combined EML
combCEML	combined complex EML
CRLB	Cramér-Rao lower bound
DOA	direction of arrival
EASI	equivariant adaptive separation via independence
ECG	electrocardiogram
EML	extended maximum likelihood

(*Acronyms*, continued)

EVD	eigenvalue decomposition
FECG	fetal ECG
FIR	finite impulse response
flop	floating-point operation
FOC	fourth-order centroid
HOEVD	higher-order eigenvalue decomposition
HOS	higher-order statistics
HOSVD	higher-order singular value decomposition
ICA	independent component analysis
i.i.d.	independent and identically distributed
ISR	interference-to-signal ratio
JADE	joint approximate diagonalization of eigen-matrices
jpdf	joint probability density function
LMS	least mean squares
LS	least squares
LTI	linear time-invariant
MA	moving average
MC	Monte Carlo
MCF	Moreau's contrast function
MECG	maternal ECG
ML	maximum likelihood
MMSE	minimum mean squared error
MRANC	multi-reference adaptive noise cancellation
MSE	mean squared error
NLMS	normalized LMS
PCA	principal component analysis
pdf	probability density function
PNLMS	power-normalized LMS
PRBS	pseudorandom binary sequence

(*Acronyms*, continued)

QAM	quadrature amplitude modulation
rv	random variable
skd	source kurtosis difference
sks	source kurtosis sum
SLLN	strong law of large numbers
SNR	signal-to-noise ratio
SOK	sum of output kurtosis
SOS	second-order statistics
SVD	singular value decomposition
TOBSE	third-order blind signal estimator
TONBSE	third-order non-blind signal estimator
WP	waveform preserving

Chapter 1

Introduction

1.1 Motivation: Problem and Applications

The problem of **blind source separation (BSS)** consists of the recovery of a set of unobservable signals — so-called *sources* — only from the observation of a set of sensor outputs which are generated as mixtures of the sources. Contrary to what one might believe at first glance, this is not always an ill-conditioned problem, even though little or nothing may be known about the original processes and the mixing structure. In many instances it suffices to assume the statistical independence of the source signals (which is quite plausible in practice) and other mild conditions on the mixing pattern defining the transformation from sources to sensors. As a matter of fact, this multichannel problem has become a topic of intensive research from the signal processing and artificial neural network communities since the mid-eighties [71], and it has been successfully solved through a considerable number of different approaches. BSS is the subject matter of the present investigation.

Why BSS has drawn so much research attention is not surprising, as it spans a vast number of application areas. Its potential is further substantiated by the fact that proceeding blindly often presents some advantages over more conventional techniques. A very typical example is the reconstruction of unknown transmitted signals impinging from unknown directions of arrival on an array of antennas (or microphones, in the acoustic environment), whose layout may also be unknown. This particular problem arises in areas such as radar, sonar and multi-user mobile radio communications. The unknown (by the receiver) transmitted signals constitute the source signals, whereas the signal received by each antenna of the array is a mixture of the sources. This application may also be modelled from the perspective of array signal processing as a direction of arrival (DOA) or a standard beamforming problem. However, such classical techniques require to parameterize the problem through the DOA of the sources and the array layout information [109, 112], which may not be known a priori. Deviations of the model from reality (so-called calibration errors) lead to errors in the estimated sources which are difficult to quantify or predict. In contrast, blind processing does not make such prior assumptions on the sources and the mixing structure, and so it is, by nature, insensitive to array calibration errors. As a consequence, blind beamforming may often outperform 'informed' beamforming [23].

Another common application example, of concern to speech processing, is the so-called 'cocktail party' problem. At a party, each guest can hear the speech of the speaker they want to listen to, together with other surrounding sounds such as the voice of other 'competing' speakers, music, background noise, etc. It is well known that the human brain is capable to separate all these perceived signals and to 'switch' its attention to the desired one. In cognitive psychology, the areas dedicated to the study of auditory grouping and selective attention aim to understand the psychological processes involved in this human perception task [58, 60]. In the early fifties, it was discovered that this ability entails making use of physical differences among the auditory messages (such as differences in the sex of the speaker, location of the speaker, voice intensity, etc.) in order to select the one of interest [38, 58–60]. Electronically, analogous results can be achieved by adequately processing the output signals of an array of microphones. In this case, the key distinctive feature allowing the identification of the sources is their statistical independence, together with certain conditions on the mixing configuration [128, 131]. These mixing conditions could be considered as the counterpart of the spatial-difference features aiding human beings to accomplish auditory separation.

The field of biomedical signal processing also contains applications which may be tackled using BSS techniques. One of the most relevant is the extraction of the fetal heartbeat signals — or electrocardiogram (ECG) — from recordings obtained through electrodes attached to the mother's skin. The electrodes output a mixture of mother's and fetus' heartbeat signals, as well as other disturbances such as thermal noise from electronic equipment, mains interference, maternal electromyogram (EMG) interference, etc. For diagnostic purposes, it is thus necessary to extract the fetal heart contributions to the corrupted cutaneous recordings. This task can be accomplished with the help of BSS methods [15, 86, 125].

Other applications of BSS comprise such diverse disciplines as [8, 23, 35, 37, 42, 48, 68, 71, 122, 128, 131, 132]:

- *Array signal processing*: separation and recognition of sources from unknown arrays, DOA estimation, source localization from perturbed arrays.

- *Radar and sonar*: jammer rejection, radar and sonar detection, airport surveillance, beamforming.

- *Communications*: mobile radio and regenerative satellite communications, data communications in the presence of interference and cross-coupling effects, blind code-division multiple access (blind CDMA).

- *Speech processing*: speech recorded in the presence of background noise and/or competing speakers, automatic voice recognition in noisy acoustic environments, aircraft cockpit environments.

- *Medicine*: determination of firing patterns of neural signals from EMG's measured from skin electrodes, elimination of ECG artifacts, separation and determination of brain activity sources.

- *Semiconductor manufacturing*: monitoring the status of the key process parameters (e.g., diffusion times, temperatures, gas fluxes, etc.) from the process testing data (e.g., threshold voltage, drive current, sheet resistance, parasitic capacitance, etc.)

- *Circuit testing and diagnosis*: temperature monitoring in high power-density electronic circuits, determination of randomly located input signals from measured output signals.

- *Image processing*: reconstruction and restoration of distorted images.

- *Statistics*: factor analysis, data pre-processing before Bayesian detection and classification, data mining.

In all the above applications, the main aim is to extract an inaccessible set of signals (the sources) from another set of measurements which originated as mixtures of the first set. When no assumptions are made on the mixing configuration and the source signals, the problem is tagged 'blind'. As quoted earlier, blind processing may prove more convenient than other more parametric approaches in many cases. In other instances, a blind approach may be the only reasonable option to face the problem, due to the lack of available information.

1.2 Thesis Objectives, Structure and Contributions

Although the treatment of non-linear mixtures has recently started to receive attention [35], this study is aimed at the context in which the mixtures can be considered as linear. Among the linear mixtures, two different types exist, namely, convolutive and instantaneous. In the former, the channel which models the transfer from each source to each sensor can be represented as a linear time-invariant (LTI) filter. In the latter, this transfer is reduced to a scalar factor, so that the whole mixing process can be represented by a simple matrix multiplication. The instantaneous linear case is specially relevant in narrowband signal processing environments, as well as in many other applications.

To solve the blind separation of instantaneous linear mixtures many different approaches exist, which will be exhaustively reviewed in the next chapter. The focus of this thesis is on the methods which offer analytic or closed-form solutions. In the basic scenario composed of two sources and two sensors, and in the absence of any noise, it has been proved that the source extraction and mixing matrix identification can be accomplished in an analytic manner [41, 42, 44, 52, 69]. Several of these apparently different solutions exist, but a closer examination reveals that they are actually interrelated. In addition, no detailed account of their properties and performance has been given to date. Hence, it is the **objective** of this thesis to consider these methods, to put forward new procedures, to disclose their links, and to carry out a comprehensive study of their properties and performance.

Below is a summary of the main original **contributions** of this work, together with its **organization** on a chapter-by-chapter basis.

Chapter 3:

- An *approximate maximum-likelihood (AML)* estimator, which was originally derived for a considerably restricted range of source distributions [69], is generalized to virtually any source probability density function (pdf) through a new expression called *extended ML (EML)*.

- The asymptotic performance analysis of the new closed-form method is carried out, resulting in an analytical expression for the sampling-distribution asymptotic pdf of the estimator. It is proved, and empirically endorsed, that for high enough sample size such pdf reduces to a Gaussian distribution. This analysis presents the additional feature that it is also applicable, under minor alterations, to the other estimators considered later in this work.

- A link is found between the closed-form formula of [41] (*Comon's formula, CF*) and the previous two methods. With the aid of the new analysis tools developed, the CF is investigated, revealing interesting new findings about its performance.

Chapter 4:

- The EML is proven to come from the optimization of a certain contrast (or cost) function. In the two-signal case, this function is later shown to loosen the applicability conditions of another contrast function previously proposed in the literature [94, 95], and to simplify its respective analytic solution given in [52]. A small-error analysis carried out on the contrast function verifies the asymptotic variance obtained from the Gaussian approximation to the EML asymptotic pdf.

Chapter 5:

- The performance of the last three closed-form methods (AML, EML and CF) is shown to depend on certain statistics of the source signals. This drawback is mitigated by proposing an alternative estimator, the *alternative EML (AEML)*, which does not exhibit such dependence. Its asymptotic performance is studied and its associated contrast function obtained.

- A combined estimation strategy is then suggested, whereby the EML or the AEML are selected according to certain decision rule. A practical empirically-based suboptimal decision rule is derived, and shown to be very close to the theoretical asymptotically optimal rule in most cases. The derivation of the sub-optimal rule is based on a novel result whereby signals of arbitrary kurtosis can be simulated by means of pseudorandom binary sequences of suitable probability (Appendix B).

Chapter 6:

- The new methods are extended to the general BSS scenario of more than two signals via a pairwise strategy [48]. Their performance is then evaluated in the context of a biomedical problem: the fetal ECG extraction. Real ECG recordings from a pregnant woman are used to assess the new methods and to compare them to other well-established BSS procedures [23, 48] and a classical technique [129].

Chapter 7:

- We capitalize on the simple yet powerful algebraic formalism of the EML, the AEML and the combined method to develop adaptive (or on-line) versions, characterized by a high convergence speed and stable convergence to valid separation solutions under mild conditions. These adaptive procedures are studied and contrasted with other adaptive methods.

Chapter 8:

- All the previous methods are based, implicitly or explicitly, on the fourth-order cumulants of the data. A general expression based on the nth-order cumulants is derived for the estimation of the pertinent parameter. The EML estimator is found to be a particular case of such general class when $n = 4$.

- At $n = 3$, the class yields a novel third-order closed-form expression, the *third-order blind signal estimator (TOBSE)*, which is analyzed and compared to the 4th-order estimators.

- An analogous general class is derived for the complex-mixture case. The use of the *bicomplex numbers*, which are defined in compliance with the structure and operations of the remaining mixing transformation after second-order pre-whitening, simplifies this development. In turn, this result highlights the unforeseen remarkable existing parallelism with the real case, and leads to a unified formulation of the analytical solutions to the real-mixture and the complex-mixture problems.

Most of the experimental methodology adopted herein is based on Monte Carlo simulations over synthetic signals of various distributions. This permits the evaluation and comparison of the methods' performance over a number of well-known distributions for the source signals. This widely-spread methodology is also useful in validating the theoretical results. Since the population (or true) values of the statistical parameters of the signals employed are known in advance, the empirical outcomes can be numerically contrasted with the expected theoretical results, providing an indication of their validity.

The above chapters embody the contributing part of this study. In addition, Chapter 2 condenses the necessary background on the BSS problem, while Chapter 9 summarizes the work, draws its conclusions and proposes possible avenues of further research. For the sake of clarity in the presentation, some information has been moved to the appendices.

1.3 Publications Derived from this Work

The following publications have arisen from the present work:

Conference papers:

- E. Bacharakis, A. K. Nandi and V. Zarzoso, "Foetal ECG Extraction Using Blind Source Separation Methods", in: *Proceedings EUSIPCO-96, VIII European Signal Processing Conference*, Trieste, Italy, September 10–13, 1996, Vol. I, pp. 395–398.

- A. K. Nandi and V. Zarzoso, "Foetal ECG Separation", in: *IEE Colloquium on the Use of Model Based Digital Signal Processing Techniques in the Analysis of Biomedical Signals*, Savoy Place, London, April 16, 1997, No. 9, pp. 8/1–8/6.

- V. Zarzoso and A. K. Nandi, "The Potential of Decorrelation in Blind Separation of Sources Based on Cumulants", in: *Proceedings ECSAP-97, First European Conference on Signal Analysis and Prediction*, Prague, Czech Republic, June 24–27, 1997, pp. 293–296.

- V. Zarzoso and A. K. Nandi, "Generalization of a Maximum-Likelihood Approach to Blind Source Separation", in: *Proceedings EUSIPCO-98, IX European Signal Processing Conference*, Rhodes, Greece, September 8–11, 1998, Vol. IV, pp. 2069–2072.

- V. Zarzoso y A. K. Nandi, "Extracción del ECG fetal mediante técnicas de separación ciega de fuentes", en: *Actas CASEIB-98, XVI Congreso Anual de la Sociedad Española de Ingenieros Biomédicos*, Valencia, 21–23 de Septiembre de 1998, pp. 89–92 [Spanish].

- V. Zarzoso and A. K. Nandi, "Source Distribution Independent Adaptive Blind Signal Separation", in: *Proceedings ICA'99, First International Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, January 11–15, 1999, pp. 65–70.

- V. Zarzoso and A. K. Nandi, "Blind Source Separation without Optimization Criteria?", in: *Proceedings ICASSP-99, IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, Arizona, USA, March 15–19, 1999, Vol. III, pp. 1453–1456.

- V. Zarzoso and A. K. Nandi, "Comparison between Blind Separation and Adaptive Noise Cancellation Techniques for Fetal Electrocardiogram Extraction", in: *IEE Colloquium on Medical Applications of Signal Processing*, Savoy Place, London, October 6, 1999, No. 107, pp. 1/1–1/6.

Refereed journal papers:

- A. K. Nandi and V. Zarzoso, "Fourth-Order Cumulant Based Blind Source Separation", *IEEE Signal Processing Letters*, Vol. 3, No. 12, December 1996, pp. 312–314.

- V. Zarzoso, A. K. Nandi and E. Bacharakis, "Maternal and Foetal ECG Separation Using Blind Source Separation Methods", *IMA Journal of Mathematics Applied in Medicine & Biology*, Vol. 14, No. 3, 1997, pp. 207–225.
- V. Zarzoso and A. K. Nandi, "Modelling Signals of Arbitrary Kurtosis for Testing BSS Methods", *IEE Electronics Letters*, Vol. 34, No. 1, January 8, 1998, pp. 29–30. (Errata: Vol. 34, No. 7, April 2, 1998, p. 703.)
- V. Zarzoso and A. K. Nandi, "Blind Separation of Independent Sources for Virtually Any Source Probability Density Function", *IEEE Transactions on Signal Processing*, Vol. 47, No. 9, September 1999, pp. 2419–2432.
- V. Zarzoso and A. K. Nandi, "Adaptive Blind Source Separation for Virtually Any Source Probability Density Function", accepted for publication in *IEEE Transactions on Signal Processing*, 1999.

Book chapter:

- V. Zarzoso and A. K. Nandi, "Blind Source Separation", in: A. K. Nandi (Ed.), *Blind Estimation Using Higher-Order Statistics*, Boston: Kluwer Academic Publishers, 1999, pp. 167–252.

Under review:

- V. Zarzoso and A. K. Nandi, "Non-Invasive Fetal Electrocardiogram Extraction: Blind Separation vs. Adaptive Noise Cancellation", submitted to *IEEE Transactions on Biomedical Engineering*, May 1999.
- V. Zarzoso and A. K. Nandi, "Analysis and Development of Closed-Form Estimators for Blind Separation of Sources — Part I: Real Mixtures", submitted to *IEEE Transactions on Signal Processing*, June 1999.
- V. Zarzoso and A. K. Nandi, "Analysis and Development of Closed-Form Estimators for Blind Separation of Sources — Part II: Complex Mixtures", submitted to *IEEE Transactions on Signal Processing*, June 1999.

Chapter 2

Blind Source Separation

This chapter unfolds the essential background on the topic of BSS, emphasis being laid on the separation of instantaneous linear mixtures. The problem is mathematically posed in Section 2.1, which also summarizes its basic assumptions and the main difficulties concerning its solution. BSS is approached from several different perspectives, allowing a more insightful clearer picture of the matter. In Section 2.2 the problem is solved by means of a two-step scheme, which is a strategy commonly adopted in the literature. Section 2.3 addresses the relevant notion of equivariance in source separation. Several indices to quantitatively measure the separation performance are presented in Section 2.4. Section 2.5 gives an exhaustive account on the different methods suggested in the literature to solve the BSS problem. Solutions based on closed-form expressions, which constitute the core of the present research, are recalled in Section 2.6.

2.1 Problem Description

Consider the set of q **source signals** $\{x_i(k) \in \mathbb{C},\ i = 1, 2, \ldots, q\}$ which generate the set of p **observations** or **measurements** $\{y_i(k) \in \mathbb{C},\ i = 1, 2, \ldots, p\}$ at the sensor output, where k represents a time index (which will be dropped in the sequel when convenient). In the general case, the measurements can be regarded as mixtures of transformed versions of the sources, contaminated by some additive **noise**

$\{n_i(k) \in \mathbb{C},\ i = 1, 2, \ldots, p\}$:

$$y_i(k) = \sum_{j=1}^{q} H_{ij}\{x_j(k)\} + n_i(k), \qquad i = 1, 2, \ldots, p, \tag{2.1}$$

where $H_{ij}\{\cdot\}\colon \mathbb{C} \mapsto \mathbb{C}$ denotes the transformation undergone by jth source contributing to the ith sensor signal. In the linear case, $H_{ij}\{\cdot\}$ is a linear time-invariant (LTI) system operating over $x_j(k)$, $H_{ij}\{x_j(k)\} = h_{ij}(k) * x_j(k)$, $h_{ij}(k) \in \mathbb{C}$, symbol $*$ standing for the convolution operator. Then, the ith measurement becomes:

$$y_i(k) = \sum_{j=1}^{q} h_{ij}(k) * x_j(k) + n_i(k), \qquad i = 1, 2, \ldots, p. \tag{2.2}$$

This is called the **convolutive mixture** BSS model [110, 128, 131, 132].[1] Abusing notation, as in [50, 62, 82], this model can also be written in matrix form as:

$$\begin{bmatrix} y_1(k) \\ \vdots \\ y_p(k) \end{bmatrix} = \begin{bmatrix} h_{11}(k) & \ldots & h_{1q}(k) \\ \vdots & \ddots & \vdots \\ h_{p1}(k) & \ldots & h_{pq}(k) \end{bmatrix} * \begin{bmatrix} x_1(k) \\ \vdots \\ x_q(k) \end{bmatrix} + \begin{bmatrix} n_1(k) \\ \vdots \\ n_p(k) \end{bmatrix} \Rightarrow$$

$$\Rightarrow \quad \mathbf{y}(k) = H(k) * \mathbf{x}(k) + \mathbf{n}(k), \tag{2.3}$$

where the column vectors $\mathbf{y} = [y_1, \ldots, y_p]^{\mathrm{T}}$, $\mathbf{x} = [x_1, \ldots, x_q]^{\mathrm{T}}$ and $\mathbf{n} = [n_1, \ldots, n_p]^{\mathrm{T}}$ contain the observations, the sources and the noise signals, respectively. Symbol $^{\mathrm{T}}$ denotes the transpose operation. Matrix H represents the multi-input multi-output linear system (channel) which links sources with measurements. The matrix notation stresses the multichannel nature of the problem. In the nomenclature of the blind deconvolution problem [138], eqn. (2.3) corresponds to a general multichannel blind deconvolution model [50]. Figure 2.1 graphically depicts the BSS model for convolutive mixtures, in the simplified scenario of two sources and two measurements.

In many applications this model can be further simplified by assuming that the linear filters h_{ij} are composed of a single coefficient: $h_{ij}(k) = m_{ij}\delta(k)$. This occurs, for instance, in narrowband propagation conditions, where the propagation delays from sources to sensors can be reduced to phase delays. As a result, the propagation effects

[1] See also [42, 47] for its moving-average (MA) parameterization.

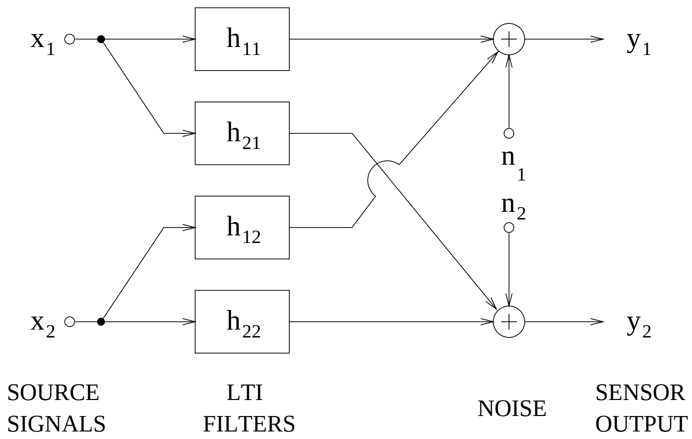


Figure 2.1: Convolutive signal separation model in the two-source two-sensor case.

can be modelled by an amplitude factor, $|m_{ij}|$, and a phase term, $e^{\jmath\angle m_{ij}}$. Under these circumstances, equation (2.2) becomes:

$$y_i(k) = \sum_{j=1}^{q} m_{ij}x_j(k) + n_i(k), \qquad i = 1,\, 2,\, \ldots,\, p, \tag{2.4}$$

or, in matrix form:

$$\begin{bmatrix} y_1(k) \\ \vdots \\ y_p(k) \end{bmatrix} = \begin{bmatrix} m_{11} & \ldots & m_{1q} \\ \vdots & \ddots & \vdots \\ m_{p1} & \ldots & m_{pq} \end{bmatrix} \begin{bmatrix} x_1(k) \\ \vdots \\ x_q(k) \end{bmatrix} + \begin{bmatrix} n_1(k) \\ \vdots \\ n_p(k) \end{bmatrix} \quad \Rightarrow$$

$$\Rightarrow \quad \mathbf{y}(k) = M\mathbf{x}(k) + \mathbf{n}(k). \tag{2.5}$$

This is the **instantaneous linear mixture** BSS model. Now matrix $M \in \mathbb{C}^{p\times q}$ contains the coefficients of the linear transformation which represents the transfer function from sources to observations. This matrix is denoted as **mixing**, **transfer** or **channel-parameter matrix** [122–124]. Figure 2.2 shows the instantaneous linear mixture model for the blind separation of two sources from two sensor outputs.

The instantaneous linear model also applies when convolutive mixtures are considered, but in a different domain [61, 79, 80]. Effectively, Fourier-transforming both sides of (2.3) yields in the frequency domain:

$$\mathbf{Y}(\omega) = \mathbf{H}(\omega)\mathbf{X}(\omega) + \mathbf{N}(\omega), \tag{2.6}$$

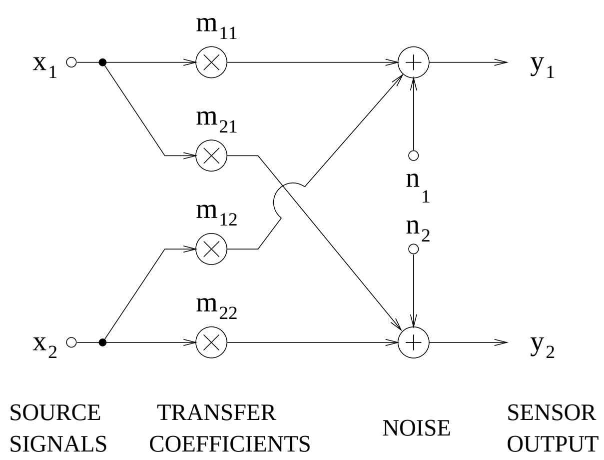


Figure 2.2: Instantaneous linear mixture model for two sources and two measurements.

where $\mathbf{Y}(\omega) = \mathrm{FT}\{\mathbf{y}(k)\}$, $\mathbf{H}(\omega) = \mathrm{FT}\{H(k)\}$, $\mathbf{X}(\omega) = \mathrm{FT}\{\mathbf{x}(k)\}$ and $\mathbf{N}(\omega) = \mathrm{FT}\{\mathbf{n}(k)\}$ are the Fourier transforms of the measurement vector, the multichannel filter, the source vector and the noise vector, respectively. Therefore, both models (2.3) and (2.5) are equivalent when contemplated from different domains. Note, however, that in (2.6) the coefficients of the corresponding transfer matrix are not necessarily constant as in (2.5), but they may well vary with ω. References [110, 128, 131, 132] address in detail the convolutive mixture case.

We will focus our attention on instantaneous linear mixtures. In this context, the **goal** of BSS can be stated as follows: to extract or recover the unknown original source signals $\mathbf{x}$ and possibly the unknown coefficients of the linear transformation that form the mixing matrix M from the only knowledge of the measured signals at the sensor output $\mathbf{y}$. The first objective (so-called *blind signal estimation* [124]) involves, explicitly or implicitly, the estimation of the transfer matrix (so-called *blind channel identification* [124]), so both goals are inextricably intertwined. Figure 2.3 graphically illustrates the basic BSS set-up.

2.1.1 Why 'blind'?

The term 'blind' emphasizes the fact that no a priori assumptions on the structure of the mixing matrix and the source signals are made, in contrast to ordinary array processing techniques. For instance, in DOA estimation the transfer matrix is parameterized as a

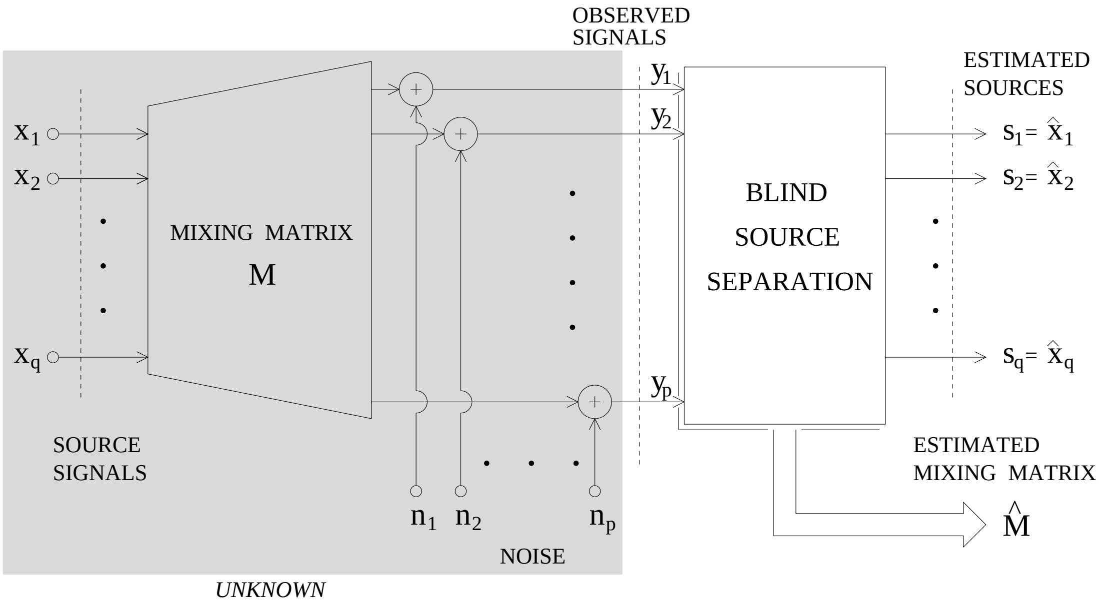


Figure 2.3: General blind source separation set-up.

function of the arrival angles and the array layout [109,112]. No such parameterization is done in BSS problems. And precisely there lies the versatility of the BSS model, since in many cases modelling the transfer between sources and sensors is very involved, or simply infeasible when no a priori information is available about the mixture. In 'informed' array processing, deviations of the parameterized model from the actual array layout (calibration errors) lead to errors in the estimated sources which sometimes are very difficult to quantify. No such deviations exist in the BSS model simply because nothing is assumed about the mixing structure. In the context of the beamforming problem, for instance, more advantages of blind techniques over 'informed' techniques are pointed out in [23]. We anticipate that the lack of information of the powerful blind operation style is compensated by means of a strong but often physically plausible hypothesis on the source signals: their statistical independence.

2.1.2 Statistical notation

Before resuming our review of the BSS problem, let us first set the statistical nomenclature and notation conventions. This is an important task, since the problem is deeply embedded in statistical principles, and *all* approaches to its solution exploit, explicitly or not, the statistical information contained in the measured data [34]. The following statistical terms will hold in the sequel (Appendix A). Given the real-valued q-component random variable $\mathbf{z} = [z_1, \ldots, z_q]^{\mathrm{T}} \in \mathbb{R}^q$, its nth-order moments are defined as:

$$\mu^z_{i_1 \ldots i_n} = \mathrm{E}[z_{i_1} \cdots z_{i_n}]. \tag{2.7}$$

Similarly:

$$\kappa^z_{i_1 \ldots i_n} = \mathrm{Cum}[z_{i_1}, \ldots, z_{i_n}], \tag{2.8}$$

represents the nth-order cumulant of the components of real-valued random vector $\mathbf{z}$. For the bivariate case, $\mathbf{z} = \{z_1, z_2\}$, this notation is simplified as

$$\mu^z_{\underbrace{1 \ldots 1}_{n-r} \underbrace{2 \ldots 2}_{r}} = \mathrm{E}[z_1^{n-r} z_2^r] = \mu^z_{n-r,r} \tag{2.9}$$

$$\kappa^z_{\underbrace{1 \ldots 1}_{n-r} \underbrace{2 \ldots 2}_{r}} = \mathrm{Cum}_{n-r,r}[z_1, z_2] = \kappa^z_{n-r,r}, \tag{2.10}$$

which is Kendall's convention for the two-dimensional case [120].

We call **kurtosis** of z_i its 4th-order marginal cumulant, κ^z_{iiii}. The **skewness** corresponds to the 3rd-order marginal cumulant, κ^z_{iii}. According to (2.10), in the two-dimensional case κ^z_{40} and κ^z_{04} represent the kurtosis of components z_1 and z_2, respectively, while κ^z_{30} and κ^z_{03} denote their skewness. Totally analogous terminology and notation hold for the other random variables and vectors used in this thesis just by changing the super-indices accordingly. Similar notation will be defined in Chapter 8 when dealing with complex variables.

Salient properties of moments and cumulants are summarized in Appendix A. These properties will be taken on board throughout this thesis, since the methods studied herein are based on the HOS of the data.

2.1.3 Assumptions

A problem with a model like (2.5) would certainly be ill-posed if no further assumptions were made about the characteristics of the system. These hypotheses can be divided into three categories, depending on whether they refer to the mixing matrix, the sources or the noise signals. The standard hypotheses are the following:

A1. The mixing matrix M is full column rank, with $p \geqslant q$.

A2. Each component of $\mathbf{x}(k)$, $x_i(k)$, is a stationary zero-mean stochastic process, $i = 1, 2, \ldots, q$.[2]

A3. At each time instant k, the components of $\mathbf{x}(k)$ are mutually statistically independent.

A4. The components of $\mathbf{x}(k)$ are unit-power.

A5. Each component of $\mathbf{n}(k)$, $n_i(k)$, is a stationary zero-mean process, $i = 1, 2, \ldots, p$.

A6. At each time instant k, the components of $\mathbf{n}(k)$ are mutually independent, as well as independent of the sources.

Assumption A1 states that the number of sensors must be at least equal to the number of sources. In general, one can recover, at most, as many sources as sensors. However, under certain conditions which hold in very specific problems, the extraction of more sources than sensors is indeed feasible [21,54]. Similarly, the source stationarity of Assumption A2 can sometimes be relaxed, or the lack of it can be tackled by a suitable adaptive implementation (such as those studied in Chapter 7).

Assumption A3 is the crux of the blind separation. Statistical independence is equivalent to the factorization of the source joint probability density function (jpdf) in the product of the marginal pdf's:

$$p_x(\mathbf{x}) = \prod_{i=1}^{q} p_{x_i}(x_i), \tag{2.11}$$

[2]Even though the source zero-mean hypothesis is always referred to in the literature, it seems to be taken for granted by all authors and no explanation is ever given. We now provide a straightforward justification: the source signals are always assumed zero-mean simply because a constant term (the non-zero mean factor) can be regarded as an additional source. To see this mathematically, consider a source signal x with mean $m \neq 0$. Then, it can be decomposed as $x = \tilde{x}+m$, where $\tilde{x}$ represents the non-constant (i.e., time-varying) zero-mean component of x. Since $\mathrm{E}[\tilde{x}^r m^s] = \mathrm{E}[\tilde{x}^r]m^s$, $\forall r, s \in \mathbb{N}$, it follows that $\tilde{x}$ and m are independent. Hence the non-zero constant factor is considered by the separation system as another independent source signal, the process $x = \tilde{x} + m$ being actually a mixture itself.

Non-zero mean sources could alternatively be accounted for through an extra constant source in the BSS model. However, in most cases the constant factors do not convey any meaningful information (see Section 2.1.4). In addition, all higher-order cumulants of a constant component are null, which poses identifiability problems (see Section 2.1.5). Consequently, it is wiser to adhere to the source zero-mean assumption.

where $p_x(\mathbf{x})$ is the source jpdf, and $p_{x_i}(x_i)$ the pdf of the ith source signal. The above necessary and sufficient independence condition can also be stated as a function of the cumulants,[3] which is often more tractable as a separation criterion:

$$\mathrm{Cum}[x_{i_1}, \ldots, x_{i_r}] = 0, \qquad \forall (i_1, \ldots, i_r) \text{ not all equal, } \forall r \in \mathbb{N}, \tag{2.12}$$

i.e., the source cross-cumulants are null [14, 92, 100].[4] In many applications, the phenomena generating the sources are physically independent (e.g., the mother's and fetus' heartbeat in the fetal ECG extraction problem, see Chapter 6), and physical independence of the sources is a sufficient condition for their independence in the statistical sense. However, the above condition on the source cumulants is too strong and it is often enough to constrain independence up to order r, that is, that all sth-order cross-cumulants be equal to zero, for $s \leqslant r$. Generally, it is chosen $r = 4$, as in [23, 29, 41, 48] and many other references. Third-order cumulants have not been paid attention in the literature because many practical distributions are symmetric. For these, the cumulants of such order are zero, and hence the problem becomes ill conditioned [42]. We will see in Chapter 8, however, than looking at $r = 3$ may prove profitable indeed.

The independence assumption allows the problem to be contemplated from an interesting algebraic viewpoint, as will be seen in Section 2.1.7. The foundations of such a perspective are laid down by noting that the rth-order cumulants (2.8) of a random variable form a multi-indexed array (i.e., a matrix with r dimensions):

$$\kappa^x_{(r)}(i_1, \ldots, i_r) \triangleq \kappa^x_{i_1 \ldots i_r} = \mathrm{Cum}[x_{i_1}, \ldots, x_{i_r}], \qquad 1 \leqslant i_1, \ldots, i_r \leqslant q,$$
$$\kappa^x_{(r)} \in \mathbb{R}^{\overbrace{q \times \cdots \times q}^{r}}. \tag{2.13}$$

The diagonal elements of this array ($i_1 = \cdots = i_r$) store the marginal rth-order cumulants, whereas the rth-order cross-cumulants are disposed in the off-diagonal entries. Hence, independence up to order r makes the sth-order cumulant array of the sources exhibit a diagonal structure, for all $s \leqslant r$. In particular, the second-order cumulant array is $\kappa^x_{(2)} = R_x$, the source spatial covariance matrix:

$$R_x \triangleq \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}], \tag{2.14}$$

[3]The unacquainted reader is referred to Appendix A for the necessary background on higher-order statistics.

[4]See Appendix A.3 for a proof.

symbol $^{\mathrm{H}}$ representing the complex-conjugate transpose operator. Thus, R_x is also diagonal.

Assumption A4 is only a normalization convention. Actually the power of the sources can be included in the columns of M, leaving the observations unaltered. This property lies in the multiplicative-group structure of the model, which makes left-multiplication of the sources by a matrix equivalent to right-multiplication of the mixing matrix by the same linear transformation (associativity) [29,63]. Let P_i be the power of the ith source, $1 \leqslant i \leqslant q$. If $\mathbf{x}$ has not unit-power signals, the equivalent set of sources

$$\tilde{\mathbf{x}} = R_x^{-\frac{1}{2}}\mathbf{x}, \qquad R_x^{-\frac{1}{2}} = \mathrm{diag}\left(\frac{1}{\sqrt{P_1}}, \ldots, \frac{1}{\sqrt{P_q}}\right), \tag{2.15}$$

is such that $R_{\tilde{x}} \triangleq \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{H}}] = I_q$, $\;q \times q$ identity matrix. Then the new equivalent mixing matrix is transformed according to:

$$M\mathbf{x} = M(R_x^{\frac{1}{2}}\tilde{\mathbf{x}}) = (MR_x^{\frac{1}{2}})\,\tilde{\mathbf{x}}, \tag{2.16}$$

so one can always think of the sources being unit-power by properly scaling the columns of the mixing matrix. Hence we can assume that

$$R_x = I_q \tag{2.17}$$

with no loss of generality.

2.1.4 Waveform-preserving solutions

The source recovery and mixing matrix identification can be accomplished up to some extent. The following technical difficulties concerning the determination of the sources must be noticed [23,122]:

1. Scalar factors can be exchanged between each source and the corresponding column of M without modifying the observations at all. That is, even bearing in mind the restriction imposed by Assumption A4, the sign (in the real case) or phase (in the complex case) of each source signal remains unobservable.

2. The outputs of the separator cannot be ordered. As a matter of fact, the order of the sources is itself arbitrary, since arranging the sources and the columns of

the mixing matrix accordingly leave the observations unchanged. It follows that the source signals can at best be recovered up to a permutation.

However, these degrees of freedom in the sources that can be estimated at the output of any unmixing system do not actually represent a limitation in the separation process, because in most practical applications:

1. The information conveyed by the sources is contained in their waveforms, rather than in their magnitudes, and hence a finite scalar factor is admissible.

2. The order in which the signals are arranged is immaterial. An order change in the recovered sources relative to the original set is thus acceptable.

Therefore, it is acceptable to obtain at the **separator output** a solution close to:

$$\hat{\mathbf{x}} = \mathbf{s} = PD\mathbf{x}, \tag{2.18}$$

where P is a $(q \times q)$ permutation matrix and D a non-singular diagonal matrix of the same size. The latter matrix represents the source re-arrangement commented in the first point above, whereas the former is a mathematical representation of the scaling factors referred to in the second place. The set of solutions $(MD^{-1}P^{-1},\, PD\mathbf{x})$ form an equivalence class, so-called **waveform-preserving (WP)** equivalence [122–124]. Hence, solutions $(M,\, \mathbf{x})$ and $(MD^{-1}P^{-1},\, PD\mathbf{x})$ are said to be equivalent in the WP sense. Any recovered source vector of the form $PD\mathbf{x}$ is called a **signal copy** of the sources [34, 39].

The extraction of the sources is usually carried out by estimating a linear transformation W such that

$$\mathbf{s} = W\mathbf{y} \tag{2.19}$$

is WP equivalent to the original set of sources $\mathbf{x}$, i.e., is a signal copy of the original sources. A matrix W fulfilling this condition is referred to as a (valid) **separating matrix**. The **global system**

$$G = WM \tag{2.20}$$

is the matrix which links the original set of sources with the estimated set, that is, the matrix that represents the whole mixing-unmixing system. Hypothesis A4 restricts

D to have unit-norm diagonal elements. The product PD is then called **quasiidentity** matrix [29]. As a conclusion, in the noiseless case the global matrix exhibits a quasiidentity structure for any valid separating matrix.

2.1.5 General approach. PCA and ICA. Identifiability

Most BSS methods turn the sources independence assumption A3 into the separation criterion. A linear transformation W is sought so that the signals at the output of the unmixing system, $\mathbf{s}$, are independent. The independence hypothesis imposes certain conditions on the source cross-cumulants, given by (2.12). Thus, the cumulants become a natural quantitative measure of the degree of independence among the components of the separator output $\mathbf{s}$. Specifically, the elements of W must be such that the cross-cumulants of $\mathbf{s}$ vanish. The order up to which these cross-cumulants must be close to zero directly depends on the order of independence assumed for the sources.

Most BSS methods mainly benefit from the spatial diversity of the problem, that is, the fact that distinct sensors output different (linearly independent) mixtures of the source signals, and generally any time structure is overlooked. Sometimes the time coherence cannot be exploited, as when independent and identically distributed (i.i.d.) observations are processed. In this connection, the BSS problem reduces to the identification of the source-vector probability distribution from the sample distribution of the observation vector [32].

Two major groups of methods for BSS can be distinguished, depending on the statistical information employed — order of cumulants and degree of statistical information of the sources exploited. The first group is composed of second-order techniques, also known as **Principal Component Analysis (PCA)**. PCA methods seek the removal of second-order dependence from the set of observations, thus providing a set of uncorrelated signals at the unmixing system output. The other main group of BSS methods is referred to as **Independent Component Analysis (ICA)**. They aim to look for an output vector whose components are higher-order independent, the term 'higher' meaning more than second order. PCA procedures are simpler than ICA methods, but they are not always able to preserve the source waveforms. The ICA overcomes this indeterminacy by resorting to the higher-order statistical information of the measurements. The use of this extra information makes the WP separation possible (with the limitations stated below). In many occasions, the BSS problem is tackled in two steps, the first one being a PCA and the second implementing ICA. This two-stage scheme presents some advantages relative to a single-step approach, as will be highlighted in Section 2.2.

The question that naturally arises at this point is: is it sufficient to constrain independence at the separator output to obtain an effective estimation of the sources, even though the time information is ignored or cannot be used? The answer is given in [124]. A necessary and sufficient condition for the blind identifiability of the the pair $(M, \mathbf{x})$ using HOS is condensed by the following theorem:

Theorem (blind identifiability [124]). *The source signals and the channel parameter matrix can be uniquely identified up to the waveform-preserving equivalence, if and only if at most one of the components of the source signals is Gaussian.*

An analogous result is reached in [48]. This identifiability condition stems from the Gaussianity property of cumulants (Appendix A.2), which states that the cumulants of order higher than two of a random variable with Gaussian distribution are all null. This bounds the applicability of cumulants of such orders in the separation of Gaussian signals. The above identifiability condition could also be deduced from the fact that any linear combination of normal random variables is also normally distributed, and hence there is no way of separating such Gaussian signals in an additive model like (2.5). The previous result is a general condition for the identifiability of instantaneous linear mixtures. Actually, particular methods may have their own specific identifiability conditions, but all of them revolve around this generic result.

In some cases the time information can also be exploited, which may offer some advantages. For instance, if the sources have different spectral content, the separation can be achieved by means of the second-order statistics of the observation signals (such as correlation at different time lags). This allows Gaussian sources to be retrieved as well [10, 122, 124]. In this thesis, however, no temporal information will be regarded.

2.1.6 Geometrical insights

Let us look at the BSS problem from a geometrical point of view. Each observation sample $\mathbf{y}$ may be represented as a point or vector in a p-dimensional vector space. If $\mathbf{m}_j$, $j = 1, \ldots, q$, denote the column vectors of the mixing matrix, then model (2.5) may alternatively be expressed in the noiseless case as

$$\mathbf{y} = \sum_{j=1}^{q} \mathbf{m}_j x_j. \tag{2.21}$$

Hence, the contribution of source x_j to the measurement vector is given by the projection of its amplitude along the direction of the vector $\mathbf{m}_j$. On this account, the columns of the mixing matrix are also termed **source directions**, **transfer vectors** or **source signatures** [17]. The measurement vector $\mathbf{y}$ is built up of the vector sum of the sources projected on the source directions. The source directions span the range or column space of M which, since it contains all possible measurement vectors generated from the source signals via (2.21), is also known as the source **signal subspace** [111]. Component-wise, source x_j is present in the ith measurement in a factor of $m_{ij}x_j$, which is the projection of $\mathbf{m}_j x_j$ on the ith-axis. Figure 2.4 depicts this notion in the 2-source 3-sensor case. The source waveforms can then be recovered by projecting back the measurement vector onto the source directions. The optimal projection is non-orthogonal (*oblique*) if the transfer vectors are not orthogonal too. Ideally, q vectors $\{\mathbf{w}_i \in \mathbb{C}^p,\ i = 1, \ldots, q\}$ are sought such that

$$\hat{x}_i = \mathbf{w}_i^{\mathrm{T}} \mathbf{y} = \sum_{j=1}^{p} \mathbf{w}_i^{\mathrm{T}} \mathbf{m}_j x_j \tag{2.22}$$

is equal to the original source x_i. For this to hold:

$$\mathbf{w}_i^{\mathrm{T}} \mathbf{m}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases} \tag{2.23}$$

Obviously, the separating matrix $W = [\mathbf{w}_1, \ldots, \mathbf{w}_q]^{\mathrm{T}}$ defined by the last equation is the pseudoinverse of the mixing matrix, $M^{\dagger}$, which unfortunately is unknown. Hence, as commented at the end of Section 2.1, the recovery of the sources is intimately related to the identification of the transfer matrix. A *spatial filtering* process is taking place when performing the separation, since each estimated source is obtained through a particular linear combination of the sensor output:

$$\hat{x}_i = \sum_{j=1}^{p} w_{ij} y_j, \tag{2.24}$$

in which $W = (w_{ij})_{q \times p}$.

2.1.7 Algebraic standpoint: cumulant-tensor diagonalization

The BSS problem may also be approached from an algebraic perspective. The cumulant array (2.13), at any order, is actually a **tensor** as it fulfils the **multilinearity property** (Appendix A.2). Given an rth-order array $\kappa^x_{j_1 \ldots j_r}$ defined on a q-dimensional coordinate system, it is said to exhibit the mentioned property when, if vector $\mathbf{x}$ is linearly transformed according to $\mathbf{y} = M\mathbf{x}$, then a similar (multi-)linear relationship holds between the associated arrays, namely:

$$\kappa^y_{i_1 \ldots i_r} = \sum_{j_1 \ldots j_r} m_{i_1 j_1} \ldots m_{i_r j_r} \kappa^x_{j_1 \ldots j_r}. \tag{2.25}$$

The cumulant arrays do so [14], and hence can be considered as tensors [91].

In algebraic terms the source extraction and transfer matrix identification are shown to be equivalent to the diagonalization of the observation cumulant tensors [49, 124]. Intuitively, this is fairly reasonable, since the source signals being independent, their cumulant tensors possess such a diagonal structure, as explained in Section 2.1.3:

$$\kappa^x_{i_1 \ldots i_r} \begin{cases} \neq 0, & i_1 = \cdots = i_r, \\ = 0, & \text{otherwise}, \end{cases} \tag{2.26}$$

where $\kappa^x_{(r)}$ is defined in (2.13). Therefore, it naturally follows that one may proceed by searching for a separating matrix W such that when applied on the observation vector it results in another random vector with diagonal cumulant tensors. Usually it suffices to constrain a diagonal shape for cumulant tensors at one or two different orders only.

In the second-order case, we are aiming at the covariance matrix diagonalization. This is illustrated in Figure 2.5. The covariance matrix is always symmetric and, as such, it can be diagonalized by using several procedures. One of them is the eigenvalue decomposition (EVD), although more robust alternative techniques exist, as will be

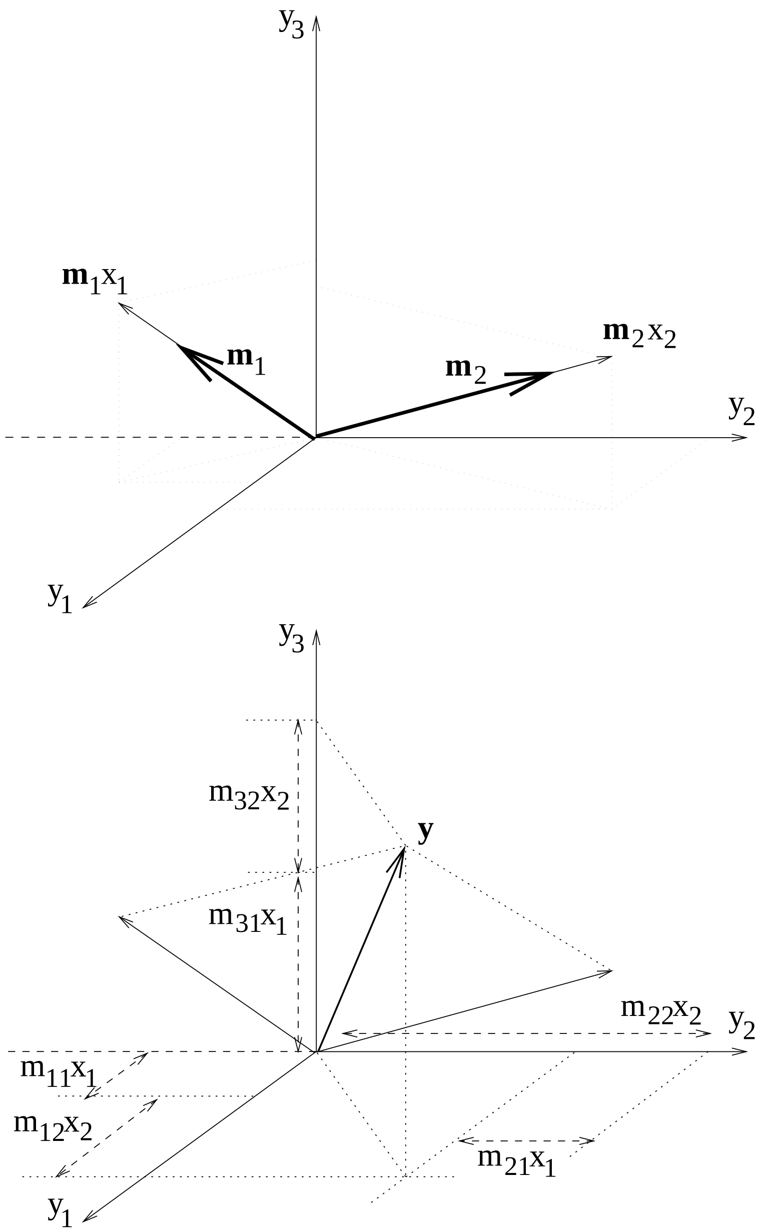


Figure 2.4: Geometrical interpretation of BSS. In the noiseless two-source three-sensor scenario, the measurement vector $\mathbf{y}$ can be represented in a three-dimensional space as the vector sum of the source amplitudes x_1 and x_2 projected along the respective source directions $\mathbf{m}_1$ and $\mathbf{m}_2$.

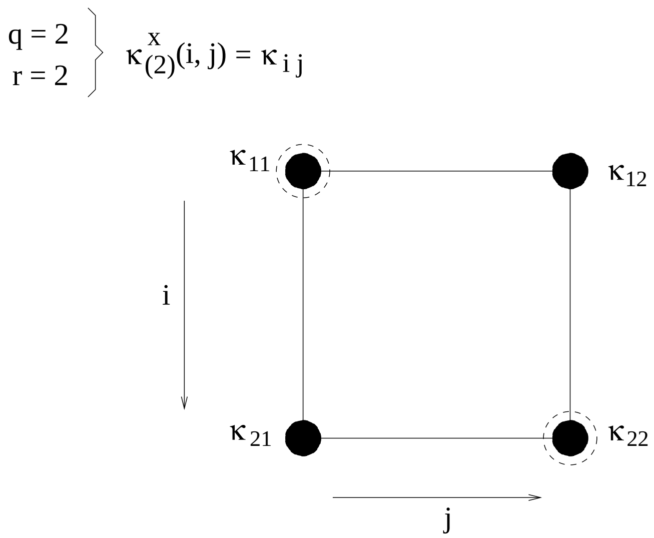


Figure 2.5: Second-order cumulant tensor structure. In the second-order case, the cumulant tensor is just the covariance matrix. Making the off-diagonal elements of such matrix zero and preserving non-null diagonal elements, the covariance matrix diagonalization is achieved, thus resulting in uncorrelated components.

discussed in the next section. The diagonalization of the covariance matrix leads to uncorrelated signals, i.e., only second-order independent (PCA of the observations).

When $r = 3$, the cumulant tensor may be represented as a cube in a three dimensional space, as shown in Figure 2.6. In the higher-order ($r > 2$) case, however, the diagonalization problem becomes more involved. Again, the higher-order cumulant tensor shows a symmetric structure, since the cumulants are symmetric function of their indices [14]. The difficulty lies in the fact that, as opposed to the second-order case where there exists a wide range of algebraic tools to deal with matrices, no such variety is available in the higher-order array domain [49]. On the other hand, the symmetric sensor-output cumulant tensor of order r and p dimensions has $\binom{p+r-1}{r}$ free parameters, introducing just as many constraints, whereas the separating matrix W is composed of qp unknowns. Consequently, only a small subset of symmetric higher-order ($r > 2$) tensors are linearly diagonalizable [49]. One is left with two choices: either to try an exact diagonalization in a space of larger dimensions ($q > p$) or to look for an approximate diagonal tensor decomposition in a space of the same dimension ($p = q$) [49]. This issue will be revisited in Section 2.5, when dealing with higher-order BSS techniques.

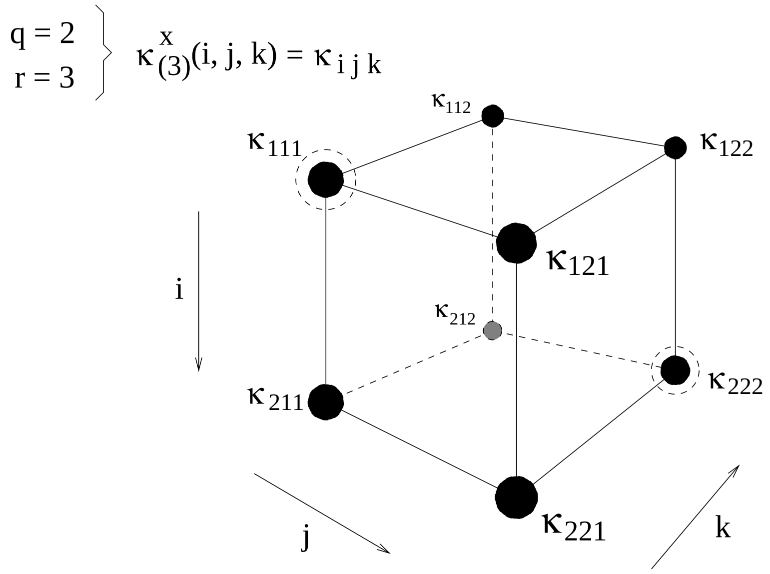


Figure 2.6: Third-order cumulant tensor structure. The elements in the diagonal (surrounded by a dashed circle) are the marginal cumulants. The off-diagonal elements represent the cross-cumulants, which must be made zero to achieve independent components. This results in a diagonal cumulant tensor.

2.2 Two-Stage Strategy

It has been shown in the literature the convenience of facing the BSS problem through a two-step strategy [2,23,48,69]. In the first step, the sensor outputs are decorrelated and normalized via conventional SOS-based techniques, which corresponds to the PCA of the observations. Afterwards, in the noiseless environment, or if the noise spatial structure is known or fulfils certain conditions, only a unitary transformation remains to be unveiled to realize the separation. When the temporal information is disregarded, the second step necessarily involves HOS, yielding the ICA of the observations. This approach presents various advantages [44, 48]: a) since SOS are generally more accurately estimated than HOS, it seems sensible to exploit first all the information they contain, and b) a dimensionality reduction is achieved, since after the SOS step fewer (less than half) mixing-matrix parameters need to be identified, which, in addition, can be attained in a very efficient manner by means of Jacobi-like iteration procedures.

Direct single-step schemes have also been devised to tackle the problem [21, 22]. They are motivated by the main drawback of the two-stage methods: the application of an initial second-order stage sets a limit on the performance of the whole separation system [24]. In addition, any existing noise affects the detection of the source signal subspace (see Section 2.2.1). If the noise is Gaussian, theoretically all their higher-order cumulant tensors vanish, so that the inaccuracies introduced by the second-order step can be avoided by exploiting the HOS directly [22]. Nevertheless, the enhanced performance is usually achieved at the expense of an increase in computational complexity.

In this work, we will adhere to the two-step design, for we are primarily interested in the study of closed-form estimators of the elementary Givens-rotation mixture which arises in the two-signal scenario after the initial second-order step.

In Section 2.2.1 the SOS-based step is reviewed. The need for further HOS-based operations is substantiated in Section 2.2.2. The discussion at the end of Section 2.2.1 leads us to concentrate our efforts into solving the noiseless BSS problem.

2.2.1 Second-order stage: PCA

The PCA is the standard statistical technique based on the decomposition of the covariance matrix of the observed variables [73, 90]. The original aim is to look for a few linear combinations of the observations that may be used to summarize the data, losing in the process as little information as possible. The selected transformation is such that it makes these linear combinations uncorrelated (i.e., independent at the second-order) and with decreasing variance. The resultant signals are regarded as an estimation of the true sources.

Let us assume that the mixing matrix has already been obtained. Then, the source signals may be estimated, for instance, via the maximum-likelihood (ML) principle by maximizing the likelihood of the observations with respect to the source vector. The log-likelihood is given by:

$$L(\mathbf{y};\mathbf{x}) = \log\big(p_y(\mathbf{y};\mathbf{x})\big) = \log\big(p_n(\mathbf{y} - M\mathbf{x})\big), \tag{2.27}$$

where $p_y(\cdot)$ and $p_n(\cdot)$ denote, respectively, the sensor-output and the noise pdf's. If the noise is Gaussian, the maximizer of (2.27) is readily determined as [111]:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \big(M^{\mathrm{T}} R_n^{-1} M\big)^{-1} M^{\mathrm{T}} R_n^{-1} \mathbf{y}, \tag{2.28}$$

where $R_n \triangleq \mathrm{E}[\mathbf{n}\mathbf{n}^{\mathrm{H}}]$ denotes the noise covariance matrix. Note that the above expression even handles the case where the noise signals are spatially correlated, provided they are Gaussian and their covariance matrix is available. On the other hand, if its known that the noise signals have the same variance, say σ_N^2, the ML estimator of the sources (2.28) reduces to:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = (M^{\mathrm{T}} M)^{-1} M^{\mathrm{T}} \mathbf{y}, \tag{2.29}$$

which corresponds to the well-known least-squares (LS) solution [111]:

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - M\mathbf{x}\|^2. \tag{2.30}$$

In the following, the singular value decomposition (SVD) [12, 66, 107] of the mixing matrix is used. It is given by:

$$M = U_{\mathrm{M}} \Sigma_{\mathrm{M}} V_{\mathrm{M}}^{\mathrm{H}}, \tag{2.31}$$

where $U_{\mathrm{M}} \in \mathbb{C}^{p\times p}$, $\Sigma_{\mathrm{M}} \in \mathbb{C}^{p\times p}$ and $V_{\mathrm{M}}^{\mathrm{H}} \in \mathbb{C}^{p\times q}$ contain, respectively, the left singular vectors, the singular values and the right singular vectors of M. Matrices U_{M} and $V_{\mathrm{M}}^{\mathrm{H}}$ fulfil the unitary condition: $U_{\mathrm{M}} U_{\mathrm{M}}^{\mathrm{H}} = U_{\mathrm{M}}^{\mathrm{H}} U_{\mathrm{M}} = V_{\mathrm{M}}^{\mathrm{H}} V_{\mathrm{M}} = I_p$.

PCA via EVD

In order to attain explicit PCA estimators of M and $\mathbf{x}$, consider the EVD of the sensor-output covariance matrix, $R_y \triangleq \mathrm{E}[\mathbf{y}\mathbf{y}^{\mathrm{H}}]$:

$$R_y = \Gamma \Lambda \Gamma^{\mathrm{H}}, \tag{2.32}$$

in which matrix Γ contains the eigenvectors of R_y and diagonal matrix Λ contains its eigenvalues. From the BSS model (2.5), R_y can also be expanded as a function of the mixture elements:

$$R_y = M R_x M^{\mathrm{H}} + R_n = M M^{\mathrm{H}} + R_n = (U_{\mathrm{M}} \Sigma_{\mathrm{M}})(U_{\mathrm{M}} \Sigma_{\mathrm{M}})^{\mathrm{H}} + \sigma_N^2 I, \tag{2.33}$$

where convention (2.17) has been borne in mind, as well as the orthogonality of the right singular matrix of M. Combining the previous two equations results in the estimate:

$$\widehat{U_{\mathrm{M}} \Sigma_{\mathrm{M}}} = \Gamma \left(\Lambda - \sigma_N^2 I\right)^{\frac{1}{2}}, \tag{2.34}$$

where the rational power is taken element-wise. From (2.29), the estimated source vector is actually the set of uncorrelated observations, so-called **whitened observations**:

$$\mathbf{z} = \left(\Lambda - \sigma_N^2 I\right)^{-\frac{1}{2}} \Gamma^{\mathrm{H}} \mathbf{y}. \tag{2.35}$$

Note, however, that with this solution the covariance matrix of the whitened output is:

$$R_z \triangleq \mathrm{E}[\mathbf{z}\mathbf{z}^{\mathrm{H}}] = \left(\Lambda - \sigma_N^2 I\right)^{-1} \Lambda \neq I. \tag{2.36}$$

Hence, if noise is present ($\sigma_N^2 \neq 0$), additional scaling is necessary to agree with convention (2.17). In the noiseless case, estimates (2.34) and (2.35) become, respectively:

$$\begin{cases} \widehat{U_{\mathrm{M}}\Sigma_{\mathrm{M}}} = \Gamma\Lambda^{\frac{1}{2}} \\ \mathbf{z} = \Lambda^{-\frac{1}{2}}\Gamma^{\mathrm{H}}\mathbf{y}. \end{cases} \tag{2.37}$$

And now $R_{\mathbf{z}} = I$. It must be stressed that $\mathbf{z}$ is not, in general, a good estimate of the source vector $\mathbf{x}$ since the right singular matrix of the mixing matrix remains unidentified. As will be seen later, only under quite specific conditions on the mixing structure can this second-order analysis provides satisfactory results.

PCA via SVD

In order to avoid the explicit computation of the observation covariance matrix, advantage can be taken of the SVD of the observation sample matrix. In a batch processing framework, T samples of the sensor signals are stored in the columns of a $p \times T$ matrix Y, so that $Y(:, k) = \mathbf{y}(k)$, where $Y(:, k)$ denotes the kth column of Y.[5] Likewise, the corresponding samples of the source and the noise signals would be contained in matrices X and N, respectively. The block-processing BSS model is the matrix equivalent of the vector model (2.5):

$$Y = MX + N. \tag{2.38}$$

The observation sample matrix Y can be expanded in its SVD:

$$Y = U\Sigma V^{\mathrm{H}}. \tag{2.39}$$

[5]This is standard MATLAB™ notation.

Then its covariance matrix accepts two different forms, one as a function of the mixing model (2.5) and the other from the previous SVD expansion:

$$R_y = MM^{\mathrm{H}} + R_n = (U_{\mathrm{M}}\Sigma_{\mathrm{M}})(U_{\mathrm{M}}\Sigma_{\mathrm{M}})^{\mathrm{H}} + \sigma_N^2 I \tag{2.40}$$

$$R_y \approx \frac{1}{T} YY^{\mathrm{H}} = \frac{1}{T} U\Sigma^2 U^{\mathrm{H}}. \tag{2.41}$$

Combining (2.40) and (2.41), one obtains the estimate:

$$\widehat{U_{\mathrm{M}}\Sigma_{\mathrm{M}}} = \frac{1}{\sqrt{T}} U \left(\Sigma^2 - T\sigma_N^2 I\right)^{\frac{1}{2}}. \tag{2.42}$$

If the noise is temporally (or spectrally) white (i.e., uncorrelated at different time instants), this expression can be substituted into (2.29), yielding the whitened signals:

$$Z = \sqrt{T}\,\left(\Sigma^2 - T\sigma_N^2 I\right)^{-\frac{1}{2}} U^{\mathrm{H}} Y = \sqrt{T}\,\left(\Sigma^2 - T\sigma_N^2 I\right)^{-\frac{1}{2}} \Sigma V^{\mathrm{H}}. \tag{2.43}$$

Again, with this solution,

$$R_z \approx \frac{1}{T} ZZ^{\mathrm{H}} = \left(\Sigma^2 - T\sigma_N^2 I\right)^{-1} \Sigma^2 \neq I, \tag{2.44}$$

and so further scaling may be necessary in order to preserve the identity structure of the source covariance matrix. That is, the signals extracted are second-order spatially white (mutually uncorrelated), due to the orthogonality of the right singular vectors of Y, but not necessarily unit variance. In the noiseless case, estimates (2.42) and (2.43) become, respectively:

$$\begin{cases} \widehat{U_{\mathrm{M}}\Sigma_{\mathrm{M}}} = \frac{1}{\sqrt{T}} U\Sigma \\ Z = \sqrt{T}\, V^{\mathrm{H}}, \end{cases} \tag{2.45}$$

where now $R_z = I$. The EVD and the SVD solution are (leaving aside round-off errors of specific implementations) equivalent. The covariance matrix factorizations (2.32) and (2.41) being equal, the relationships

$$\begin{cases} \frac{1}{\sqrt{T}} U = \Gamma \\ \Sigma^2 = \Lambda, \end{cases} \tag{2.46}$$

hold, and then (2.34) and (2.42) are the same. The difference between the two procedures lies in the fact that in many cases the covariance matrix of the observations is ill-conditioned. In such a situation, the SVD-based procedure proves more reliable, as it spares the likely errors occurred in the direct computation of the sensor-output covariance matrix.

Source-number detection and noise-variance estimation

Both EVD solutions (2.34)–(2.35) and SVD solutions (2.42)–(2.43), albeit simple and elegant, lack practical feasibility. The noise variance σ_N^2 is unknown in a genuine BSS situation. If the number of sensors is greater than the number of sources, i.e., $p > q$, σ_N^2 can be estimated from the smaller singular values of Y, or from the smallest eigenvalues of its covariance matrix. In the likely case in which these smallest singular/eigen values are not equal, their average may be taken. However, even if the exact number of sources is known, experiments show that in many cases the singular values of the noise cannot be distinguished from those of the information or desired signals [6].

In reality, the number of sources is often unavailable. Another important issue then becomes the detection of the number of sources, q, also called dimension of the signal subspace. This is closely related to the above problem, since this value needs to be known in advance to allow the estimation of the unknown noise variance and, in turn, the proper identification of the source signal subspace. Again, the detection could be done by counting the number of singular values of the observation covariance matrix (or, alternatively, of the observation sample matrix) significantly greater than zero, but experiments confirm that this procedure is not generally reliable, for the same reasons quoted in the previous paragraph. The situation is aggravated if the noise does not have the same variance or is not spatially uncorrelated,[6] but the noise covariance structure exhibits an unknown arbitrary shape. Only in the noiseless case this count yields an accurate estimate of the number of sources contained in the mixture, since the mixing matrix then spans a subspace of exactly q dimensions.

The source-subspace detection issue is discussed, e.g., in [61], where a successful novel approach is proposed for periodic narrowband sources in spatially correlated and spectral coloured noise (with shorter correlation length than the sources), based on spectral matrices of the observation vector at different time lags plus certain techniques borrowed from image processing.

[6] This would violate the first part of Assumption A6, but might well occur in practice.

Noiseless model

In conclusion, when nothing is known a priori about the noise, its effects have to be neglected and one needs to assume that the **noiseless BSS model**

$$\mathbf{y} = M\mathbf{x} \tag{2.47}$$

is being treated, where the PCA solutions are given by (2.37) and (2.45). The number of sources is hence assumed to be the same as the number of observations $q = p$.

Only if it is known that an authentic noiseless problem is being treated, further simplifications are still possible. Specifically, in such a case each of the p observed signals comes from a linear combination of q statistically independent source signals, which in particular are linearly independent (it is readily seen that uncorrelatedness implies linear independence). This means that the rank of the sample matrix Y is q, which can be obtained as the number of its singular values significantly greater than zero. As a result, Y accepts a reduced SVD expansion in the form [12, 66]:

$$Y = U_q \Sigma_q V_q^{\mathrm{H}} \tag{2.48}$$

where $U_q = U(:,\, 1\!:\!q)$, $\Sigma_q = \Sigma(1\!:\!q,\, 1\!:\!q)$ and $V_q = V(:,\, 1\!:\!q)$. Therefore the estimates in (2.45) can be computed by using these reduced matrices, and analogously for the EVD solutions (2.37).

One final comment about classical statistical nomenclature is in order. Strictly speaking, when the observations follow a noisy model such as (2.5)–(2.38), one is actually performing a **Factor Analysis (FA)** on the observed variables to obtain estimates of the mixing matrix and the source signals. Only when the noise is neglected FA and PCA become identical [73, 90].

2.2.2 Higher-order stage: ICA

Second-order analysis leaves the right singular matrix of the mixing matrix undisclosed. This implies that the whitened sensor output and the true source vector are related through a unitary transformation $Q \in \mathbb{C}^{q \times q}$:

$$\mathbf{z} = Q\mathbf{x}. \tag{2.49}$$

Remark that matrix Q above is not necessarily equal to the right singular matrix, $V_{\mathrm{M}}^{\mathrm{H}}$, of M. This can be justified as follows. Matrix $\widehat{B}^{\dagger} = \big(\widehat{U_{\mathrm{M}}\Sigma_{\mathrm{M}}}\big)^{\dagger}$ is said to whiten the observations, and is thus named the **whitening matrix**. It fulfils: $R_z = R_{\widehat{B}^{\dagger}y} = \widehat{B}^{\dagger}R_y\widehat{B}^{\dagger\mathrm{H}} = (\widehat{B}^{\dagger}U_{\mathrm{M}}\Sigma_{\mathrm{M}})(\widehat{B}^{\dagger}U_{\mathrm{M}}\Sigma_{\mathrm{M}})^{\mathrm{H}} = I_q$. Hence, the estimated whitening matrix may indeed be any of the form $\widehat{B}^{\dagger}U_{\mathrm{M}}\Sigma_{\mathrm{M}} = \tilde{Q}$, with $\tilde{Q}$ an arbitrary unitary matrix. Therefore, $\mathbf{z} = \widehat{B}^{\dagger}\mathbf{y} = Q\mathbf{x}$, as in (2.49), with $Q = \tilde{Q}V_{\mathrm{M}}^{\mathrm{H}}$.

An estimate, $\widehat{Q}$, of Q is then sought so that it provides the source signal copy:

$$\hat{\mathbf{x}} = \mathbf{s} = \widehat{Q}^{-1}\mathbf{z}, \qquad \widehat{Q} = \widehat{\mathbb{Q}}(\mathbf{z}), \tag{2.50}$$

in which $\widehat{\mathbb{Q}}(\mathbf{z})$ is a suitable estimator of the unknown unitary matrix computed from the whitened observations.

In order to unveil this transformation, more constraints are needed. If the time information is used, such transformation may be estimated from the auto-correlation matrix of the whitened processes at lags different from zero (as done, e.g., by methods such as AMUSE [122] or SOBI [10], commented in Section 2.5). If, by contrast, the temporal information cannot be exploited or is just disregarded, the estimation of this matrix requires to take further advantage of the source statistical independence assumption. After a first estimation of the sources via a 2nd-order procedure, the matrix Q is sought by constraining higher-order independence at the outputs. This process involves, explicitly or implicitly, resorting to the higher-order statistical information (higher-order cumulants) of the signals. The extra processing, the ICA of the observations, supplies a set of higher-order independent signals. Under the general identifiability conditions summarized in Section 2.1.5, these output signals are WP equivalent to the wanted sources. In conclusion, the indeterminacy vanishes with the use of HOS techniques. Figure 2.7 sketches this two-step approach to BSS. Some of the most significant HOS-based approaches to BSS are surveyed in Section 2.5.

The indeterminacy of SOS-based methods accepts another interesting interpretation. By virtue of formulae (2.34) and (2.42) [or their noiseless-model counterparts in (2.37) and (2.45)], the PCA solutions are doubly orthogonal: the estimated mixing matrix has orthogonal columns (spatial orthogonality), whereas the whitened output is composed of uncorrelated components (temporal orthogonality) [19, 39]. However, in most situations the spatial orthogonality does not occur, rendering PCA solutions

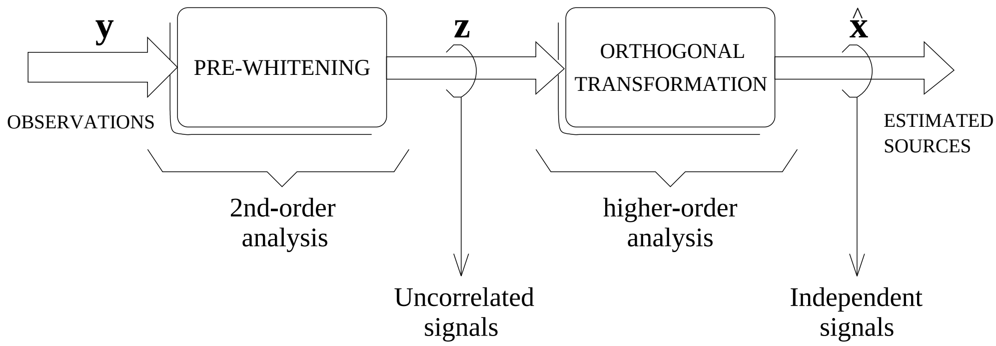


Figure 2.7: The two-stage approach to signal separation.

unacceptable. Figure 2.8 illustrates this issue in a noiseless real-mixture two-source two-sensor scenario. Each cross represents a sample of the sensor output, for which its abscissa value corresponds to the amplitude of the first sensor and its ordinate value the amplitude of the second sensor: $\big(y_1(k),\, y_2(k)\big)$. This is the two-dimensional equivalent of Figure 2.4 for specific sample values.[7] The mixing-matrix columns are chosen to be non-orthogonal, but the source directions estimated by the PCA (via SVD) are orthogonal, thus not achieving a separating solution. By contrast, the ICA step is able to identify the non-orthogonal (in the spatial sense) components of the mixture, while preserving the temporal orthogonality entailed by the source statistical independence.

The two-signal real case

In the simplified real-mixture 2-source 2-sensor BSS scenario, unitary matrix Q in (2.49) becomes an elementary planar Givens rotation of the form:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{2.51}$$

The problem, then, is reduced to determining angular parameter θ from the whitened observations. Once this has been achieved, a counter-rotation of this angle performed on the whitened observations provides the original source waveforms. The complex case is of similar nature, and will be presented when addressed in the second part of Chapter 8.

[7]This kind of representations, known as scatter diagrams, will be introduced in detail in the next chapter.

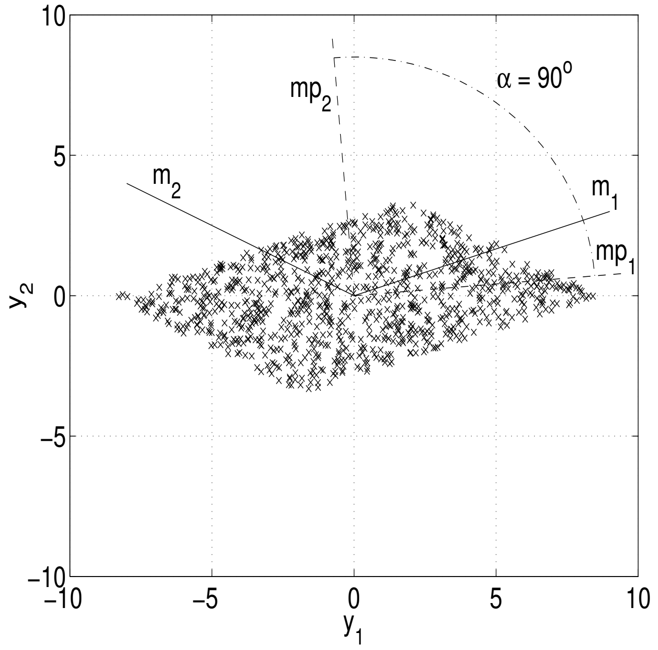


Figure 2.8: Indeterminacy of second-order analysis. The plot displays a 1000-sample scatter diagram of a mixture of two uniformly distributed sources. Vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ represent the true source directions, chosen to be non-orthogonal. Vectors $\mathbf{m}_{p_1}$ and $\mathbf{m}_{p_2}$ show the source directions obtained by the PCA (via SVD). They are orthogonal.

The characteristics of this simplified scenario are indeed restrictive. However, a clear understanding of the problem in this basic set-up is of primal importance, since the solution in more elaborate scenarios (such as that composed of more than two signals, as we will see in Chapter 6) may benefit from results on this fundamental case.

2.3 Equivariance and Uniform Performance

The multiplicative structure of the noiseless BSS model (2.47) brings into play the notion of **equivariance**. An estimator $\widehat{\mathbb{M}}$ of M is said to be equivariant when [24, 25]:

$$\widehat{\mathbb{M}}(A\mathbf{y}) = A\widehat{\mathbb{M}}(\mathbf{y})V, \qquad \forall A \text{ invertible}, \tag{2.52}$$

for some irrelevant quasiidentity matrix V. Hence, equivariance means that transformations in the observed data produce analogous transformations in the estimated parameter. Considering the global matrix obtained through an equivariant estimator of the mixing matrix $\widehat{M} = \widehat{\mathbb{M}}(\mathbf{y})$, we have:

$$\begin{aligned} G = WM &= \widehat{\mathbb{M}}(\mathbf{y})^{\dagger}M && \text{since } W = \widehat{M}^{\dagger} \\ &= \widehat{\mathbb{M}}(M\mathbf{x})^{\dagger}M && \text{by the BSS model (2.47)} \\ &= \big(M\widehat{\mathbb{M}}(\mathbf{x})V\big)^{\dagger}M && \text{by equivariance} \\ &= V^{\dagger}\widehat{\mathbb{M}}(\mathbf{x})^{\dagger}. && \end{aligned} \tag{2.53}$$

That is, the global matrix, and hence the source signals, obtained through an equivariant estimator of the mixing matrix do not depend at all on the mixing matrix itself, but only on the sources. This attractive property enjoyed by equivariant algorithms is known as **uniform performance**. An estimator fulfilling (2.52) is also referred to as *fully invariant.* Along the same lines, an estimator $\widehat{\mathbb{Q}}$ of the unitary matrix Q is called *orthogonal invariant* if

$$\widehat{\mathbb{Q}}(U\mathbf{z}) = U\widehat{\mathbb{Q}}(\mathbf{z})V, \qquad \forall U \text{ unitary}, \quad V \text{ quasiidentity}. \tag{2.54}$$

In the noiseless case, orthogonal invariant algorithms are also fully invariant [24].

The question now is if such algorithms exist and, in the case they do, how to obtain them. This is answered in [25]. It is shown that there is actually a wide variety of equivariant BSS methods. Among them, methods based on maximum-likelihood estimation as well as procedures relying on contrast functions are easily proved to be equivariant. An adaptive separation algorithm may acquire this property by means of two devices: serial updating and the relative gradient [2, 25, 27, 29, 84].

2.4 Performance Indices

In order to assess the separation quality offered by BSS methods in quantitative terms, some performance indices need to be defined. A variety of such indicators are employed throughout the literature. Herein, we will highlight five of them, which are useful in evaluating the following aspects of the separation results: a) the similarity between the waveforms of the original sources and the estimated ones, b) the resemblance between the original mixing matrix and the estimated one, c) to what extent the sources interfere in the estimation of each other, d) the degree of statistical independence (at certain order) in the recovered source signals, and e) the computational requirements of a specific method in achieving the separation. In order to obtain the first three measures, the original sources and mixing matrix must be available (as they are, e.g., in computer simulations involving synthetic signals), while the last two can be used regardless. These parameters are all examined next. Later, a commentary is made regarding the fairness of these performance measures.

2.4.1 Source-waveform preservation

To measure the resemblance between the original and the reconstructed source waveforms, one can resort to their mean squared difference, that is:

$$\varepsilon^2 \triangleq \mathrm{E}\big[\|\hat{\mathbf{x}} - \mathbf{x}\|^2\big]. \tag{2.55}$$

As the sources may be estimated up to a factor PD (Section 2.1.4), the signals in (2.55) should be properly arranged and scaled before computing ε^2. This can be automatically done, for instance, by means of the spatial cross-covariance matrix $\mathrm{Cov}[\mathbf{x}, \hat{\mathbf{x}}] = \mathrm{E}[\mathbf{x}\hat{\mathbf{x}}^{\mathrm{H}}]$, which indicates in which position of the recovered source vector the estimate of each original source signal appears. When ε^2 is normalized with respect to the source power, it is named **crosstalk** [40, 88]. Assuming that the signals are unit-variance, index ε^2 can be developed into:

$$\varepsilon^2 = \frac{1}{q}\sum_{i=1}^{q} \mathrm{E}\big[|x_i - \hat{x}_i|^2\big] = 2\left(1 - \frac{1}{q}\sum_{i=1}^{q} \mathrm{I\!Re}\big(\mathrm{E}[x_i \hat{x}_i^*]\big)\right), \tag{2.56}$$

with $\mathrm{I\!Re}(\cdot)$ standing for the real part and symbol * for complex conjugation. Hence, the better the separation, the closer this parameter is to zero, because in that case, $\hat{\mathbf{x}} \approx \mathbf{x}$ and hence $\mathrm{E}[x_i \hat{x}_i^*] \approx 1$. On the other hand, if the recovered source waveforms do not look much like the original ones, they are both uncorrelated in statistical terms, and then $\mathrm{E}[x_i \hat{x}_i^*] \approx 0$, so $\varepsilon^2 \approx 2$. Accordingly, we define the percentage of **waveform similarity** as:

$$\varepsilon^2\,(\%) = 100\big(1 - \varepsilon^2/2\big)\,\%. \tag{2.57}$$

In a logarithmic scale, ε^2 can be defined as:

$$\varepsilon^2\,(\mathrm{dB}) = 10 \log_{10} \varepsilon^2. \tag{2.58}$$

2.4.2 Quality of the estimated mixing matrix

The following indicator was introduced in [48] to measure the distance between two matrices, modulo a post-multiplicative factor of the form PD. Given two invertible matrices A and B, this index reads:

$$\begin{aligned}\epsilon(A,\, B) \triangleq \sum_i \Big| \sum_j |D_{ij}| - 1 \Big|^2 + \sum_j \Big| \sum_i |D_{ij}| - 1 \Big|^2 + \\ + \sum_i \Big| \sum_j |D_{ij}|^2 - 1 \Big| + \sum_j \Big| \sum_i |D_{ij}|^2 - 1 \Big|,\end{aligned} \tag{2.59}$$

with

$$D = \widehat{A}^{-1}\widehat{B}, \qquad \begin{cases} \widehat{A} = A\Delta_A^{-1}, & \Delta_A(k,\, k) = \|A(:,\, k)\| \\ \widehat{B} = B\Delta_B^{-1}, & \Delta_B(k,\, k) = \|B(:,\, k)\|. \end{cases} \tag{2.60}$$

Since

$$\epsilon(A,\, B) = 0 \quad \Leftrightarrow \quad B = APD, \tag{2.61}$$

by using $\epsilon(M,\, \widehat{M})$ the resemblance between the true mixing matrix M and the one estimated by the separation algorithm $\widehat{M}$ may be calculated. In an effective source extraction, $\epsilon \approx 0$. In dB:

$$\epsilon\,(\mathrm{dB}) = 10 \log_{10} \epsilon. \tag{2.62}$$

2.4.3 Interference-to-signal ratio

The parameter

$$\varrho_{ij} \triangleq \frac{(WM)_{ij}^2\, \mathrm{E}\big[|x_j|^2\big]}{(WM)_{ii}^2\, \mathrm{E}\big[|x_i|^2\big]}, \qquad i \neq j, \tag{2.63}$$

quantifies the relative power contribution of the jth source in the estimation of the ith source. It is hence called **interference-to-signal ratio (ISR)** [23, 32] obtained in rejecting the jth source in the estimate of the ith source by a separating matrix W. For this reason it is also known as **rejection rate** [24, 29], and in the terminology of channel equalization as **intersymbol interference (ISI)** [28, 29]. With the unit-variance source convention:

$$\varrho_{ij} = \frac{|g_{ij}|^2}{|g_{ii}|^2}. \tag{2.64}$$

Its mean value over the indices ij can be taken as another measure of separation quality:

$$\mathrm{ISR} = \mathop{\mathrm{E}}_{ij}[\varrho_{ij}], \tag{2.65}$$

and in dB it is defined in a similar way as ε^2 and ϵ above.

2.4.4 Degree of statistical independence

The degree of statistical independence at order r of the components of the random vector $\hat{\mathbf{x}}$ may be measured as the ratio of sum of the squares of their rth-order marginal cumulants to the sum of the squares of all rth-order cumulants:

$$\Upsilon_r \triangleq \frac{\sum_i |\kappa^{\hat{x}}_{i\ldots i}|^2}{\sum_{i_1\ldots i_r} |\kappa^{\hat{x}}_{i_1\ldots i_r}|^2}, \qquad \kappa^{\hat{x}}_{i_1\ldots i_r} \triangleq \mathrm{Cum}[\hat{x}_{i_1}, \ldots, \hat{x}_{i_r}]. \tag{2.66}$$

In compliance with Section 2.1.7, this performance index is also a measure of how close the rth-order cumulant tensor of the recovered sources is to a diagonal structure. The original sources being independent, in a successful separation Υ_r will be very close to one, since the cross-cumulants of independent signals vanish. Sometimes, specially for comparison purposes, it is more useful to consider the **distance to independence** on a logarithmic scale, that is,

$$\tilde{\Upsilon}_r\,(\mathrm{dB}) = 10\log_{10}(1 - \Upsilon_r). \tag{2.67}$$

For the same degree of statistical independence, index Υ_r depends on the relative variance of the components involved. For meaningful measures, it is then convenient to normalize the signals (e.g., to unit variance) before computing this parameter.

2.4.5 Computational complexity

The floating point operations taken by an algorithm to supply a separation solution can be regarded as a appropriate measure of its computational complexity. A floating point operation (or, abbreviated, **flop**) usually corresponds to a multiplication followed by an addition [48]. However, since multiplications are, in general, considerably more costly than additions, we will consider a flop as a real multiplication. Hence, a complex product takes 4 flops, and the multiplication of a real by a complex quantity (real-complex product) 2 flops.

2.4.6 Fairness of performance measures

Care must be taken when choosing a performance index to measure the relative performance of BSS methods. This is because in some cases the objective function optimized by a method may be directly linked to the performance measure being used. As a result, that method would unfairly yield apparently better results. For instance, the measure of independence at 4th-order Υ_4 is maximized if the sum of squared kurtosis is maximized, which is precisely the contrast function of the ICA method of [48] (see next section for a thorough description). Indeed, the degree of statistical independence is affected by the particular choice of marginal and cross cumulants, as well as by how these statistics are weighted [39, 118].

Accordingly, if the true sources and mixing matrix are available, the waveform similarity between the former and the estimated sources, the ISR, or the gap between the true and the estimated mixing matrix are all, in the above sense, more impartial measures of separation quality. However, in a practical blind problem the information required to compute such parameters is, by definition, unavailable. Hence, in many occasions one is left with a heuristic performance assessment based, for instance, on visual inspection of the obtained waveforms, together with certain considerations related to the application in hand. This matter will be illustrated in Chapter 6 in the context of the fetal ECG extraction problem.

2.5 Approaches to BSS

Having reviewed the basic foundations surrounding the problem of the blind separation of instantaneous linear mixtures, it is time to survey specific separation methods. We classify the different procedures according to the nature of their respective approach. The closed-form solutions are reserved a special separate place in Section 2.6, for they constitute the pivotal theme of our research.

2.5.1 Seminal work

BSS has been thoroughly studied since 1986. Jutten provided one of the first significant approaches to this problem [71, 74] (see also [75]). The proposed solution was based on a neural network implementation (denoted by *HJ network*), but no theoretical explanation was available at that time. The neuro-algorithm aims at the cancellation of certain non-linear odd functions of the separator outputs, which implicitly introduce higher-order moments of the output components. In [43] and [46] the algorithm is explained in statistical terms, and is found to be able to separate the sources if they have non-Gaussian and symmetric pdf's. Moreover, in some pathological cases the algorithm may fail to give a proper solution. The stability analysis of the HJ network is accomplished in [117].

2.5.2 Exhaustive search

Lacoume and Ruiz propose an exhaustive search of the absolute minima of a response function built up from the cumulants of the Fourier transform of the data signals [79]. In [82], the same researchers resort to constrained optimization (via the Fletcher-Powell algorithm). This time the cost function to minimize consists in the sum of squares of the 4th-order pairwise cross-cumulants of the outputs. The output signals must also be uncorrelated, which becomes the constraint, although no source variance normalization is imposed. For the noiseless two-source two-sensor scenario, Mansour and Jutten propose as a cost function a particular 4th-order cross-cumulant [88]. In order to avoid spurious solutions, uncorrelatedness has also to be forced at the outputs, so a system of two non-linear equations turns up. The computational burden of these optimization-based procedures is more than apparent. In a bid to relieve this complexity, it is shown in [96] that the last system can be considerably simplified and the numerical search spared by applying a previous decorrelation processing to the sensor-output signals. This result, which endorses the convenience of a two-stage approach to BSS [133], relies mainly on [41] and [42].

2.5.3 Polynomial direct solutions

For the noiseless two-source two-sensor scenario a direct solution is achieved by rooting a fourth-order polynomial equation using only fourth-order cumulants if the sources are non-Gaussian [89]. Under identical conditions, that solution is further simplified in [40], where the separating matrix is estimated by rooting only a second-degree polynomial equation involving fourth-order cumulants. Both methods are quite simple and their accuracy is only limited by the precision of the cumulant estimation. Their major drawback is that, in order to fix the source-amplitude ambiguity, procedures of this sort often normalize the separating matrix (for instance, taking its diagonal elements equal to 1). This leads to algorithms whose behaviour depend strongly on the mixing matrix, hence renouncing the uniform performance property [25].

2.5.4 Contrast function optimization: ICA

The concept of ICA was originally given a solid mathematical formulation in [45, 48]. The ICA is closely related to the concept of contrast function, which is a function of the output pdf whose maximization supplies independent components at the output of the separator.

Definition (contrast function [45, 48, 49]). Let ϵ_n be the space of random variables of dimension n admitting a density. A *contrast function* is a mapping Ψ from the set of densities $\{p_s,\ \mathbf{s} \in \epsilon_n\}$ to $\mathbb{R}$ satisfying:

1. $\Psi(p_s)$ does not change if the components s_i are permuted: $\Psi(p_{Ps}) = \Psi(p_s)$, for any permutation matrix P.

2. $\Psi(p_s)$ is invariant under scale change: $\Psi(p_{Ds}) = \Psi(p_s)$, for any invertible diagonal matrix D.

3. $\Psi(p_s)$ decreases by linear combination: if $\mathbf{x}$ has independent components, then $\Psi(p_{Ax}) \leqslant \Psi(p_x)$, for any invertible matrix A.

In addition, a contrast function is said to be *discriminating* over a set ϵ if the equality $\Psi(p_{Ax}) = \Psi(p_x)$ holds only when $A = PD$, $\mathbf{x}$ being a random vector of ϵ with independent components.

The notion of contrast function is extended to the blind deconvolution problem in [50].

The HOEVD method

A specific contrast function is proposed in [48], based on minimization of the mutual information between the components at the separator output (which is directly related to the Kullback-Leibler divergence between the output-vector pdf and its pdf if it was made of independent components).[8] After some manipulations on the Edgeworth expansion [78,120,126] of the source jpdf, the contrast function simplifies into the sum of the output squared rth-order marginal cumulants:

$$\Psi_r(\widehat{Q}) = \sum_{i=1}^{q} (\kappa^s_{\underbrace{ii\ldots i}_{r}})^2. \tag{2.68}$$

Note that the criterion depends on $\widehat{Q}$ through $\mathbf{s} = \widehat{Q}^{-1}\mathbf{z}$. At 4th-order, the above contrast becomes the sum of the output squared kurtosis:

$$\Psi_4(\widehat{Q}) = \sum_{i=1}^{q} (\kappa^s_{iiii})^2. \tag{2.69}$$

It is also proven [45,48] that in the BSS context pairwise independence is tantamount to mutual independence. Therefore, when a contrast of polynomial (and, by extension, analytic) form is considered, its pairwise (that is, treating signal pairs consecutively) optimization is equivalent to its full mutual optimization. Accordingly, one may focus on the 4th-order contrast:

$$\Psi_4(\widehat{Q}) = (\kappa^s_{40})^2 + (\kappa^s_{04})^2, \tag{2.70}$$

whose optimization is carried out through the rooting of a 4th-order polynomial, its unknown variable being the angle of the sought planar rotation Q. For more than two signals, the pairwise contrast-function optimization is accomplished in a Jacobi-like iteration, via the application of successive elementary planar Givens transformations to each whitened signal pair, in turn, over a number of sweeps. As a by-product, since function (2.70) increases monotonically after each pairwise iteration [45], the 4th-order cumulant tensor is progressively being diagonalized.[9] Hence, the procedure resembles Jacobi's technique for the diagonalization of Hermitian matrices [66,85]. On this account, we refer to the method developed in [48] as *Higher-Order EVD (HOEVD)*.

[8] These are concepts borrowed from the field of information theory [55]. An excellent, insightful review on the relevance of these information-theoretical tools in the BSS context is accomplished in [34].

[9] See Section 2.1.7 for notions on cumulant tensor diagonalization.

Remark that the above definition of contrast function is reminiscent of the optimization criteria proposed in [114] in the context of the blind deconvolution problem (see the theorem in [114, Section III]). In the blind deconvolution criteria of [114], all local extrema are shown to be valid deconvolution solutions. By contrast, to date no proof has been given of the absence of spurious local maxima in the HOEVD iterative optimization procedure.

Another form of implementing ICA may be found in [104]. Its contrast function is based on the Kullback-Leibler divergence and the kernel estimate of the source probability densities. Other cost functions derived earlier in the literature are also found to be contrasts, such as JADE's cost function [25, 32] (see next section).

2.5.5 Multilinear algebra

HOS is combined with matrix algebra by Cardoso in [17], where the separation is attained through the EVD of a matrix made up of particular fourth-order moments. This technique, the *Fourth-Order Blind Identification (FOBI)* algorithm, is able to recover the sources when they have different non-Gaussian symmetric pdf's, but fails in other cases. In [18] the same author improves the method by resorting to the fourth-order moments tensor of the observations. This second version of Cardoso's algorithm performs better for any set of non-Gaussian probability distributions.

Recall (Section 2.1.7) that the multilinearity property of cumulants allows them to be considered as tensors. This property is exploited in [19–21], where tensors are regarded as linear operators acting on matrix spaces. At 4th-order, special symmetries are unveiled (giving rise to the so-called 'supersymmetry'), so that the identification of the mixing matrix can be accomplished via particular eigen-decompositions. The uniqueness of such decompositions is guaranteed even for more sources than sensors [21]. In our opinion, this result is in apparent contradiction with the tensor-decomposition bounds established in [49] (Section 2.1.7).

Analogous algebraic ideas are exposed by De Lathauwer and colleagues. In [85] they extend the SVD to higher-order arrays — yielding the *Higher-Order SVD (HOSVD)* — and apply it to the BSS problem. The cumulant array of the whitened observations can be diagonalized with this new algebraic tool, thus providing a set of independent signals. The method tackles the fetal ECG problem with success [86]. However, we believe that the decomposition exhibits uniqueness problems for sources of the same distribution, as evidenced by a number of simulations. Further, Clarke points out [39] that the nature of the decomposition restricts the method validity to sub-Gaussian sources.

The JADE method

In [23], the maximization of the contrast function

$$\Psi_{\text{JADE}}(\widehat{Q}) = \sum_{i,\,k,\,l=1}^{q} \left|\kappa^{s}_{iikl}\right|^2, \tag{2.71}$$

which is equivalent to the minimization of the sum of the squared cross-cumulants with distinct first and second indices, is proposed to find the remaining unitary transformation. The particular benefit of this contrast lies in the fact that it is linked to underlying cumulant-tensor eigenstructures that allow for its efficient optimization by means of the so-called *joint diagonalization*.

Effectively, in [23] the eigenstructure of the cumulant tensor is investigated (along the lines of [17–22]) and the relevant fourth-order statistical information (a subset of cross-cumulants) is contained in a reduced set of matrices, referred to as *eigen-matrices*. The maximization of (2.71) is then proven to be tantamount to the simultaneous diagonalization of this set, which can be efficiently carried out by a Jacobi-type algorithm, similar to that of the HOEVD method (Section 2.5.4). The resulting method, so-called *Joint Approximate Diagonalization of Eigen-matrices (JADE)*, is successfully applied to narrowband beamforming. Interestingly, the asymptotic performance of JADE is the same as Comon's HOEVD [25].

Ideas from both the HOEVD and JADE algorithms are combined in [87], and applied to the diagonalization of third-order tensors resulting from particular contractions of the 4th-order cumulant tensor. The associated contrast function is the sum of the squares of 4th-order cumulants with no more than one distinct index.

2.5.6 Indentifiability studies

Waveform-preserving identifiability issues are discussed at length in [122] and [123]. Some identifiable BSS models (both with uncorrelated sources and with independent sources) are presented and two practical blind identification algorithms developed. The first of them is called *EFOBI (Extended FOBI)* and is a generalization of Cardoso's FOBI method (Section 2.5.5) which does not ignore the noise effects, but still it only works for sources with different kurtosis values. The EFOBI is further extended to scenarios with spatially correlated noise in [116]. The other method — *Algorithm for Multiple Unknown Signals Extraction (AMUSE)* — exploits the second-order statistics of the source processes, when they are not spectrally white. It consists of the diagonalization of the whitened covariance matrix at a time lag distinct from zero. In order for this method to work properly, the sources must possess different autocorrelation at some time lag, i.e., different spectral content.

In [124], Tong *et al.* prove that the minimum-variance unbiased estimates of the source signals can be obtained if and only if the transmission channel (the transfer matrix) can be blindly identified. In turn, the channel can be blindly identified if and only if there is no more than one Gaussian source. Nevertheless, if the temporal information of the observed processes is employed, the channel may be identifiable even if the sources are all Gaussian when the second-order statistics of the observation signals (such as correlation at different times) are used. In addition, the problem of channel identification is shown to be equivalent, from an algebraic point of view, to simultaneously diagonalizing the observation cumulant matrices (unfolded representations of the corresponding cumulant arrays).

In [16] a general approach to the problem is given, which provides a uniform framework for the fundamental theory and covers the existing results as special cases. Two important concepts in the BSS problem are studied: separability and separation principles. *Separability* refers to the conditions that the measured signals have to satisfy for the sources to be identifiable. It turns out that separability depends only on the structure of the mixing matrix. *Separation principles* correspond to the conditions that can be applied to verify whether a successful separation has been carried out. Basically, if the signals at the output of the separator are pairwise independent, they must be the sources, as long as there are at most one Gaussian source signal. This outcome entirely agrees with identifiability conditions developed earlier.

2.5.7 Maximum likelihood

A ML approach is undertaken in [62]. With the help of the Gram-Charlier expansion [78, 120, 126] of the sources jpdf, a suboptimal two-step procedure is developed: first, decorrelation; second, 4th-order independence. Interestingly, the cost function to be maximized in the second step is identical to the contrast function derived separately by Comon in [48]. In [103], the ML principle leads to estimating equations which depend on non-linear functions of the observations. The optimal non-linearities are the source score functions, i.e., the derivatives of the logarithm of the source pdf's. The ML standpoint is also adopted in [69], where by using certain simplifications a closed-form solution for the angle to be estimated in the noiseless two-source two-sensor BSS set-up is found (Section 2.6). In [9] the ML principle is combined with the EM (Expectation-Maximization) algorithm to yield a method able to separate more sources than sensors in the case of discrete sources.

2.5.8 Infomax

An apparently new information-theory based approach attracted a lot of interest, the so-called *information-maximization (infomax)* principle [8]. In short, the application of this principle to source separation consists in maximizing a particular contrast function: an output entropy. The procedure is readily extended to the blind deconvolution problem. However, it is proved in [31] that, regarding source separation, infomax is equivalent to ML.

2.5.9 Adaptive separation

The topic of adaptive source separation is addressed at length by Cardoso and Laheld in [29]. They derive the *Equivariant Adaptive Separation via Independence (EASI)* algorithms, possessing the desirable uniform performance feature (Section 2.3). Other adaptive BSS methods appear in [41] and [68]. The former — *adaptive rotation (AROT)* — is the adaptive version of the block method described therein and in [42], and essentially reduces to an adaptive cumulant estimation. The latter procedure is an adaptive implementation of Gaeta-Lacoume's [62] and Comon's [48] block methods. EASI and AROT, as well as other techniques, will be considered in more detail when studying adaptive separation methods in Chapter 7.

2.5.10 Artificial neural networks

Many adaptive separation procedures accept, or can be reformulated as, specific *artificial neural network (ANN)* implementations. Typical ANN models for BSS include feedback (recurrent) or feedforward-feedback architectures [2,76]. The HOS are implicitly generated by means of suitable non-linearities in the separation network. However, in the BSS context the basic statistical principles behind these ANN realizations [2] remain the same as in non-neural implementations.

2.5.11 Signal modelling for performance testing

Many procedures (adaptive or otherwise) for BSS resort to the 4th-order statistical information of the observed data to achieve the separation, as in [29,41,48,69,88], etc. As a consequence, their performance is somehow dependent on the source statistics at that order. In [135] an interesting result is derived and its effects investigated: a simple binary distribution exhibits all possible range of normalized kurtosis values when the probability of its two events is varied accordingly. This outcome suggests that any signal may be substituted, as far as its kurtosis value and other relevant statistics are concerned, for binary distributions of appropriate parameters, which has favourable implications when testing the performance of BSS methods as a function of such statistics. This hypothesis is successfully checked out with the adaptive method of [29]. For the sake of completeness, these results are included in Appendix B, as they will be exploited at several stages of this investigation.

2.5.12 Only SOS

Along the lines of the JADE (Section 2.5.5) and AMUSE (Section 2.5.6) algorithms, and based on the identifiability criteria developed in [123] and [124], a *Second-Order Blind Identification (SOBI)* algorithm is reported in [10]. It exploits the time coherence of the source signals and it consists of a joint diagonalization of several whitened covariance matrices at different time lags. Robustness is significantly increased at low additional cost by processing such a matrix set rather than a unique matrix as in AMUSE. Like this latter algorithm, SOBI allows — in contrast to higher-order cumulant techniques — the separation of Gaussian sources. Nevertheless, the spectral separation of the source signals is still essential for a satisfactory performance of the method.

2.5.13 Convolutive and non-linear mixtures

For the more general case in which the mixing system is formed by unknown LTI filters (Section 2.1), signal separation criteria are developed in [110,128,131,132]. The solutions proposed in [110] rely on SOS only, and on the parameterization (both MA and ARMA) of the separation structure. Two types of methods are suggested. The first type is based on the minimization of the sum over different lags of squared cross-correlations of the separator output. The second type adopts a system identification approach, using the so-called recursive prediction error method. In [128], only the second-order independence (uncorrelatedness) of the sources is exploited through a decorrelation-based technique, whereas [131] is based on HOS via the polyspectra [92, 99,100]. It is proved again that statistical independence between the source signals is a sufficient condition for signal separation, but even weaker sufficient conditions involving their cross-polyspectra are found. A unified framework for many of these methods is presented in [132], which also analyzes their statistical stability for finite sample size.

State-of-the-art results on convolutive as well as on non-linear mixtures can be found in [35]. Nonetheless, in the present investigation our attention is exclusively focused on the instantaneous linear mixture problem.

2.6 Closed-Form Solutions

Diverse authors have developed analytical solutions for the noiseless two-signal case. These solutions provide explicit estimators, with no search involved, for the unknown parameters which define the unitary matrix Q relating the true sources to the signals after pre-whitening [eqns. (2.49) and (2.51)]. The first of such solutions is given for the real-mixture problem in [41] and [42]. The formula depends on the 4th-order cross-cumulants of the data, and reads:

$$\mathrm{tg}\,\hat{\theta} = -\frac{\rho}{2} + \mathrm{sign}(\rho)\sqrt{\frac{\rho^2}{4}+1}, \qquad \rho \triangleq \frac{\kappa_{31}^z - \kappa_{13}^z}{\kappa_{22}^z}. \tag{2.72}$$

The extension of this result to complex mixtures appears in [44], [53] and [80].

Harroy and Lacoume [69] apply the ML principle together with the Gram-Charlier expansion of the source jpdf. Under some extra assumptions, another closed-form expression for the estimation of the orthogonal transformation is derived:

$$\mathrm{tg}4\hat{\theta} = \frac{\sum_k r_k^4 \sin(4\phi_k)}{\sum_k r_k^4 \cos(4\phi_k)}, \qquad z_1(k) + j z_2(k) = r_k \mathrm{e}^{j\phi_k}, \quad j^2 \triangleq -1, \tag{2.73}$$

which can also be written as a function of the 4th-order statistics of the data.

In [39], Clarke derives a very similar expression by resorting to intuitive geometric arguments when the source pdf's are symmetric, although no sound statistical interpretation is given:

$$\hat{\theta} = \frac{1}{4}\arg\Big(\sum_k r_k^4 \mathrm{e}^{j4\phi_k}\Big), \tag{2.74}$$

where the "arg$(\cdot)$" function provides the angle (argument) of its complex argument with respect to the real axis, in the range $[-\pi,\, \pi]$.

For infinite sample size all the above methods (under their respective validity conditions) provide identical results, the true value of the sought parameter. The motivation for considering such a variety of direct procedures lies in that their finite-sample performances, which is what one would obtain in practice, are potentially different, even in the absence of any noise. Hence, the statistical characterization of these methods is indeed a relevant issue. Nevertheless, to date no study of the relationships and properties of these estimators has been carried out. This study is precisely one of the main goals of this thesis.

In Chapter 3 the connections among these estimators are explored, and the properties and performance of the methods analyzed. It also seen that the performance of such estimators suffers a significant degradation when the source kurtosis sum is close to zero. Motivated by this finding, in Chapter 5 another estimator is put forward which depends on the source kurtosis difference. A combined estimation strategy is then proposed, with a practical (suboptimal) decision rule derived from a number of experiments. Extensions to the general BSS scenario composed of more than two sources and observations, with application to an exciting biomedical problem, are carried out in Chapter 6, whereas adaptive implementations are examined in Chapter 7.

Another analytic solution is found in [52] for the maximization of the contrast function of [94, 95] in the two-signal case. This contrast requires the sign of the 4th-order marginal cumulants of the source to be equal and known. In Chapter 4 the mentioned requirements are spared, so that the applicability of such contrast is broadened while its analytic solution gets simplified. Connections with the ML principle are also highlighted.

Finally, Chapter 8 generalizes this class of 4th-order estimators. A general formula is found for the estimation of the unknown parameter through the nth-order statistics (cumulants) of the data. This discovery leads to a new closed-form 3rd-order estimation expression for the real case. Some of these results are also spanned to the complex-mixture environment in Chapter 8, leading to a coherent formulation of the real and complex BSS problems within the framework of their analytical solutions.

In conclusion, the present work aims to conceive an innovative unifying wider perception of the closed-form estimators for BSS.

2.7 Summary

This chapter has recalled the central aspects of the BSS problem, concentrating on the separation of instantaneous linear mixtures, and has made a thorough review of the literature on the subject. In the case no time information is used, a two-step approach has been advocated, aiming at second-order and higher-order independence, respectively. In the noiseless case, a unitary transformation remains to be estimated after the first step to complete the source extraction and mixing matrix identification. Closed-form solutions have been developed for the estimation of this matrix in the two-signal environment. The purpose of this thesis is to study such methods and develop new procedures along similar lines.

Next chapter initiates this investigation by looking into the connections among the real-mixture estimators set out in Section 2.6, and analyzing their statistical properties and performance.

Chapter 3

Closed-Form Solutions for Real Mixtures

3.1 Introduction

Having presented the essential background on the BSS problem, we are now ready to go into details. We start our investigation by addressing the direct solutions in the real-mixture scenario when a block of signal samples are processed (batch or block treatment). The basic elements peculiar to the BSS environment of real-valued signals and mixtures are presented in Section 3.2. Section 3.3 proposes a new 4th-order estimator, the so-called *extended ML (EML)*, which generalizes Harroy-Lacoume's *approximate ML (AML)* procedure [69]. By virtue of this generalization, we prove that the AML can actually be used outside the restrictions originally stated in its definition. As a matter of fact, the EML is applicable as long as the source kurtosis sum is not zero, overriding all the previous limitations of the AML. The new estimator is also linked to Clarke's method, and interpreted from an insightful geometrical standpoint. The EML statistical properties and performance are exhaustively analyzed. Section 3.4 is devoted to Comon's formula [41, 42]. Its connection to the AML and the EML is established, and its performance is understood in detail through the corresponding asymptotic analysis. In summary, in this chapter we intend to make the first approach to the direct solutions to the real-valued BSS problem from a unified perspective.

This chapter is the most dense of the present thesis. Much of the material in the following chapters relies on the exposition herein, and hence we need to be specially thorough when presenting these results. As a by-product, this thoroughness will allow convenient simplifications when reaching other results in the next chapters.

3.2 Preliminaries

This section sets the grounds for the forthcoming development. Section 3.2.1 poses the specific problem we aim to solve in the real-mixture BSS scenario. The particular relationships between the source and the whitened-signal statistics are summarized in Section 3.2.2. Finally, Section 3.2.3 reviews the so-called scatter diagram. Algebraically, this simple tool is extremely useful in simplifying certain parts of the mathematical development and, geometrically, it provides insightful interpretations for certain source distributions.

3.2.1 The real-mixture problem after pre-whitening

As explained in the previous chapter (Section 2.2), after pre-whitening, the sources and the whitened vectors are related through a unitary transformation. In the case of real mixtures, and in the simplified scenario of two sources and two sensors, such transformation becomes an elementary Givens rotation matrix:

$$\mathbf{z} = Q\mathbf{x}, \tag{3.1}$$

with

$$Q = \mathcal{Q}(\theta), \tag{3.2}$$

where

$$\mathcal{Q}(\cdot) : \mathbb{R} \mapsto \mathbb{R}^{2\times 2}, \qquad \mathcal{Q}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{3.3}$$

Evidently, angle θ characterizing the rotation is unknown. Therefore, the identification of this matrix, and thus the source separation, is reduced to the estimation of a single relevant parameter, angle θ, from the whitened observations $\mathbf{z}$. If $\hat{\theta}$ represents a valid estimate of this parameter, the sources are recovered through:

$$\hat{\mathbf{x}} = \mathbf{s} = \widehat{Q}^{\mathrm{T}}\mathbf{z}, \qquad \widehat{Q} \triangleq \mathcal{Q}(\hat{\theta}). \tag{3.4}$$

Equivalence class of valid solutions

Due to the allowed indeterminacies commented in Section 2.1.4, not only the true value of θ provides a valid separation solution. Indeed, any $\hat{\theta}$ such that

$$\mathcal{Q}(-\hat{\theta})\mathcal{Q}(\theta) = PD \tag{3.5}$$

with P a permutation and D a regular diagonal matrix, accomplishes a successful WP separation. Accordingly, the following solutions are said to be WP equivalent to the true value θ, i.e., $\hat{\theta} \equiv \theta$:

$$\hat{\theta} = \theta + \frac{n\pi}{2}, \qquad \forall n \in \mathbb{Z}. \tag{3.6}$$

Effectively, for any of the above solutions the estimated orthogonal rotation becomes:

$$\widehat{Q} = \begin{bmatrix} \cos(\theta + n\frac{\pi}{2}) & -\sin(\theta + n\frac{\pi}{2}) \\ \sin(\theta + n\frac{\pi}{2}) & \cos(\theta + n\frac{\pi}{2}) \end{bmatrix}, \tag{3.7}$$

which, expanding the trigonometric functions, may be expressed in the form:

$$\widehat{Q} = QF, \qquad F = \begin{bmatrix} \cos(n\frac{\pi}{2}) & -\sin(n\frac{\pi}{2}) \\ \sin(n\frac{\pi}{2}) & \cos(n\frac{\pi}{2}) \end{bmatrix}. \tag{3.8}$$

Matrix F is a quasiidentity matrix for all values of n, which leads to a global matrix with the same structure when counter-rotation $\widehat{Q}^{\mathrm{T}}$ is applied over the whitened observations, as in (3.4). In consequence, we call

$$[\theta] = \{\hat{\theta} \in \mathbb{R} \;:\; \hat{\theta} \equiv \theta\} \tag{3.9}$$

the *equivalence class of valid solutions* to the real-valued two-signal problem. Matrix Q can then be restricted to the set of inner rotations.

In more intuitive terms, if a rotation of arbitrary angle θ has been performed on the source jpdf, any counter-rotation of angle $\theta \pm \frac{\pi}{2}n$, for $n = 0, 1, 2, \ldots$, provides the source components, possibly with an axis interchange and sign variation, effects that are irrelevant as far as the separation is concerned. This is why all possible values of θ can be reduced to the interval $[-\pi/4, \pi/4]$, yet still providing acceptable separation solutions.

3.2.2 Statistical relationships

The output of the second-order pre-whitening stage is a set of zero-mean unit-variance uncorrelated components. In such a case, special links hold for the HOS of the data. In particular, it will be useful to recall the following expressions for the pairwise 4th-order cumulants [120]:

$$\begin{aligned} &\kappa_{40}^z = \mu_{40}^z - 3, && \kappa_{31}^z = \mu_{31}^z, && \kappa_{22}^z = \mu_{22}^z - 1 \\ &\kappa_{04}^z = \mu_{04}^z - 3, && \kappa_{13}^z = \mu_{13}^z && \\ &\kappa_{31}^x = \kappa_{13}^x = \kappa_{22}^x = 0. && && \end{aligned} \tag{3.10}$$

The first two lines express the 4th-order cumulants of the whitened measurements as a function of their moments. Totally analogous expressions hold for the kurtosis and cross-cumulants of the sources. The final identity (zero source 4th-order cross-cumulants) comes from the source independence assumption. Owing to the unit-variance convention, all statistics become normalized or standardized [48].

From eqns. (3.1) and (3.2), and using the multilinearity property (Appendix A.2), the 4th-order cumulants of the whitened observations can be expanded as a function of the unknown angle θ and the 4th-order cumulants of the independent sources:

$$\kappa_{40}^z = \cos^4\theta\,\kappa_{40}^x + \sin^4\theta\,\kappa_{04}^x \tag{3.11a}$$

$$\kappa_{31}^z = \cos^3\theta\,\sin\theta\,\kappa_{40}^x - \cos\theta\,\sin^3\theta\,\kappa_{04}^x \tag{3.11b}$$

$$\kappa_{22}^z = \cos^2\theta\,\sin^2\theta\,(\kappa_{40}^x + \kappa_{04}^x) \tag{3.11c}$$

$$\kappa_{13}^z = \cos\theta\,\sin^3\theta\,\kappa_{40}^x - \cos^3\theta\,\sin\theta\,\kappa_{04}^x \tag{3.11d}$$

$$\kappa_{04}^z = \sin^4\theta\,\kappa_{40}^x + \cos^4\theta\,\kappa_{04}^x \tag{3.11e}$$

symbols κ_{40}^x and κ_{04}^x denoting the 4th-order marginal cumulants, or kurtosis, of the source signals.

3.2.3 The scatter diagram

Consider the pair of signals $\mathbf{z}(k) = [z_1(k),\, z_2(k)]^{\mathrm{T}}$. The bidimensional plot of the amplitude points $\big(z_1(k),\, z_2(k)\big)$, $k = 1, 2, \ldots,$ is called *scatter diagram* [13, 39, 69]. The scatter diagram is an approximation of the true jpdf of the corresponding pair of variables. Areas within the scatter diagram of high point density correspond to a high value

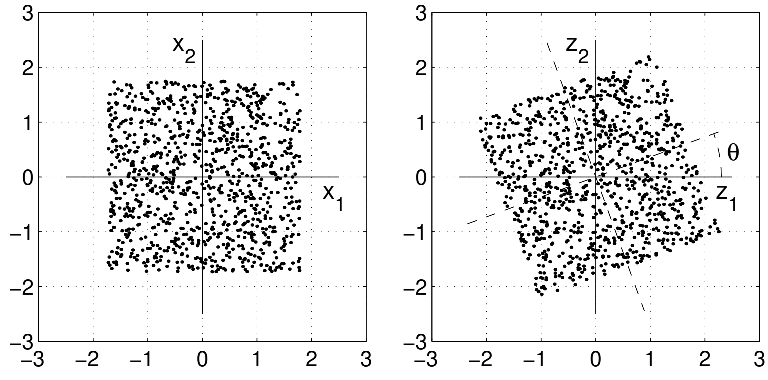


Figure 3.1: Scatter diagram for 1000 samples of: (a) original sources, two uniformly distributed independent signals; (b) sensor signals after pre-whitening. The latter corresponds to the former rotated by an unknown angle θ.

for the associated jpdf around the same area. The left panel of Figure 3.1(a) shows the scatter diagram of 1000 realizations of two uniformly distributed independent random variables (x_1, x_2), the source signals. It can be appreciated, in the first place, how the density is, roughly, uniform over the definition range, which means that the jpdf is flat. Also, it corresponds to a jpdf of independent variables, since it can be decomposed as the product of its two one-dimensional marginal pdf's. The effect of an orthogonal transformation is a scatter-plot rotation. Effectively, the scatter diagram corresponding to the whitened pair of variables (z_1, z_2), connected to the sources according to expression (3.1), is represented in the right-hand panel of Fig. 3.1. It preserves the exact shape of the source scatter-plot, but it is rotated an angle θ with respect to the new axes. Observe that the source directions are the symmetry axes of the pre-whitened jpdf, represented by the dashed lines in Fig. 3.1(b). From this perspective, source separation would be reduced to finding these symmetry axes and, by means of a suitable counter-rotation of angle θ, aligning them with the signal axes [13, 39]. As easily illustrated with the aid of Fig. 3.1, additional rotations of integer multiples of $\pi/2$ rad are also acceptable, giving rise to the equivalence class of valid solutions introduced in Section 3.2.1.

Mathematically, the scatter diagram points accept a polar as well as a complex form representation:

$$\left.\begin{array}{l}\big(x_1(k),\, x_2(k)\big) = x_1(k) + jx_2(k) = r_k e^{j\phi'_k} = r_k\angle\phi'_k \\ \big(z_1(k),\, z_2(k)\big) = z_1(k) + jz_2(k) = r_k e^{j\phi_k} = r_k\angle\phi_k\end{array}\right\} \quad k = 1,\, 2,\, \ldots, \quad j^2 \triangleq -1, \tag{3.12}$$

where, from (3.1) and (3.2)–(3.3), the angles ϕ'_k and ϕ_k are readily related by

$$\phi_k = \phi'_k + \theta. \tag{3.13}$$

Equivalently, we can write:

$$z_1 + jz_2 = (x_1 + jx_2)e^{j\theta}. \tag{3.14}$$

These polar and complex form representations will prove very convenient at certain points of the development, and will be helpful in providing a very interesting geometrical interpretation of the estimator introduced in the next section.

3.3 Generalization of an Approximate Maximum-Likelihood Method

3.3.1 An approximate maximum-likelihood (AML) estimator

By definition, the ML estimator of θ is the value of the rotation angle that maximizes the (log-)likelihood of the given whitened observations, $\mathbf{z}(k)$. Mathematically, this idea can be expressed as:

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\hat{\theta}} \log p_{\mathbf{z}}(\mathbf{z}\,|\,\hat{\theta}). \tag{3.15}$$

$p_{\mathbf{z}}(\cdot)$ denotes the pdf of the decorrelated measurements, which, due to unitary transformations (3.1) and (3.4), is related to the source pdf through:

$$p_{\mathbf{z}}(\mathbf{z}\,|\,\hat{\theta}) = p_{\mathbf{x}}(\widehat{Q}^{-1}\mathbf{z}) = \prod_{i=1}^{q} p_{x_i}\big((\widehat{Q}^{\mathrm{T}}\mathbf{z})_i\big), \tag{3.16}$$

$p_{x_i}(\cdot)$ being the marginal pdf of the ith source. For T independent observations $\mathbf{z}(k)$, $k = 1,\, 2,\, \ldots,\, T$, the overall log-likelihood becomes the sum of each sample log-likelihood:

$$\log p_{\mathbf{z}}(\mathbf{z}\,|\,\hat{\theta}) = \sum_{k=1}^{T} \log p_{\mathbf{z}}\big(\mathbf{z}(k)\,|\,\hat{\theta}\big). \tag{3.17}$$

In [69], the source jpdf was approximated by its Gram-Charlier expansion [78, 126] truncated at the fourth-order term (hence the adjective 'approximate'). In order for this expansion to be a valid pdf approximation, the kurtosis of each source must lie in the range $[0, 4]$, which excludes all short-tailed (platykurtic, sub-Gaussian, with negative kurtosis) pdf's and even some long-tailed (leptokurtic, super-Gaussian, with positive kurtosis) ones, such as the exponential distribution. For symmetric sources, i.e., with zero skewness ($\kappa_{30}^x = \kappa_{03}^x = 0$),[1] the log-likelihood function reads:

$$\log p_{\mathbf{z}}\big(\mathbf{z}(k) \,|\, \hat{\theta}\big) = -\frac{r_k^2}{2} + \frac{3}{4!}\left(\frac{r_k^4}{8} - r_k^2 + 1\right)(\kappa_{40}^x + \kappa_{04}^x)$$
$$+ \frac{1}{4!\,2}(r_k^4 - 6r_k^2)\cos 2(\phi_k - \hat{\theta})(\kappa_{40}^x - \kappa_{04}^x) + \frac{r_k^4}{4!\,8}\cos 4(\phi_k - \hat{\theta})(\kappa_{40}^x + \kappa_{04}^x). \quad (3.18)$$

Assuming, additionally, that the sources have the same distribution (at least up to the fourth-order in their Gram-Charlier expansion), the maximization of the likelihood function provides the AML estimator of θ:

$$\hat{\theta}_{\mathrm{AML}} = \frac{1}{4}\,\mathrm{arctg}\frac{\sum_k r_k^4 \sin 4\phi_k}{\sum_k r_k^4 \cos 4\phi_k}. \quad (3.19)$$

The properties of the estimator were also studied in [69]. Under all the above assumptions, it was found to be unbiased for any sample size. In addition, the Cramér-Rao lower bound (CRLB) for the estimation of θ under the conditions of this particular ML development was deduced to be:

$$\mathrm{Var}[\hat{\theta}_{\mathrm{AML}}] \geqslant \frac{6}{T[(\kappa_{40}^x)^2 + (\kappa_{04}^x)^2]}, \quad (3.20)$$

where $\hat{\theta}$ is expressed in radians.

The same authors realized in [69] the lack of sensitivity of the estimator (3.19) to the data pdf, testing its robustness for sources out of the validity domain of its development, such as uniformly distributed signals, whose kurtosis is negative. Surprisingly, the results obtained were very favourable. An explanation for this unexpected phenomenon was given for sources preserving jpdf symmetry from a geometrical standpoint based on the scatter plot. We carried out further experiments involving asymmetric sources with kurtosis values outside the interval $[0, 4]$ (e.g., exponential distributions). Positive results were also obtained. All these empirical outcomes aroused the suspicion that

[1] Refer to Appendix A.1.4 for an explanation on the concepts of skewness, kurtosis and tailness.

the validity scope of expression (3.19) was actually less restricted than originally predicted. In fact, it is proved in the next section that the AML estimator admits quite a straightforward generalization whereby almost any source joint distribution can be treated.

3.3.2 Generalization: the Extended ML Estimator

Dropping the time index k in the sequel for convenience, let us define the 4th-order complex statistic:

$$\xi \triangleq \mathrm{E}[r^4 \mathrm{e}^{j4\phi}] = \mathrm{E}[(z_1 + jz_2)^4]. \tag{3.21}$$

According to relationships (3.12) and (3.13), equation (3.21) accepts an expansion as a function of the sources and the unknown rotation angle:

$$\xi = \mathrm{e}^{j4\theta}\mathrm{E}[r^4 \mathrm{e}^{j4\phi'}] = \mathrm{e}^{j4\theta}\mathrm{E}[(x_1 + jx_2)^4]. \tag{3.22}$$

But, from the expressions given in (3.10),

$$\mathrm{E}[(x_1 + jx_2)^4] = \sum_{r=0}^{4} \mathrm{E}[x_1^r (jx_2)^{4-r}] = \kappa_{40}^x + \kappa_{04}^x, \tag{3.23}$$

and so ξ may be expressed as a function of the source statistics:

$$\xi = \mathrm{e}^{j4\theta}(\kappa_{40}^x + \kappa_{04}^x). \tag{3.24}$$

The term in brackets, the *source kurtosis sum (sks)*, is not known, because by definition the sources are not known either. However, from (3.10) and (3.12) this statistic can be obtained from the whitened data as follows:

$$\begin{aligned} \gamma \triangleq \kappa_{40}^x + \kappa_{04}^x &\underset{\substack{\uparrow \\ \text{from (3.23)}}}{=} \mathrm{E}[(x_1 + jx_2)^4] = \mathrm{E}[(x_1^4 + x_2^4 - 6x_1^2x_2^2) + j4(x_1^3x_2 - x_1x_2^3)] \\ &\underset{\substack{\uparrow \\ \kappa_{31}^x = \kappa_{13}^x = 0}}{=} \mathrm{E}[x_1^4 + x_2^4 - 6x_1^2x_2^2] = \mathrm{E}[(x_1^2 + x_2^2)^2 - 8x_1^2x_2^2] \underset{\substack{\uparrow \\ \mathrm{E}[x_1^2x_2^2]=1}}{=} \mathrm{E}[(x_1^2 + x_2^2)^2] - 8. \end{aligned} \tag{3.25}$$

That is:

$$\gamma = \mathrm{E}[(x_1^2 + x_2^2)^2] - 8 = \mathrm{E}[r^4] - 8. \tag{3.26}$$

Hence, this parameter is also available as a function of the whitened data as:

$$\gamma = \mathrm{E}[(z_1^2 + z_2^2)^2] - 8, \tag{3.27}$$

since, from (3.12), $r^2 = x_1^2 + x_2^2 = z_1^2 + z_2^2$. Therefore, from (3.24) and γ we obtain:

$$\angle \xi_\gamma \equiv 4\theta, \qquad \xi_\gamma \triangleq \xi \cdot \mathrm{sign}(\gamma). \tag{3.28}$$

In practice, for single signal realizations composed of T samples, the ensemble or population averages in (3.21) and (3.26)–(3.27) are estimated by their unbiased time or sample averages (ergodicity is implicitly assumed):

$$\hat{\xi} = \frac{1}{T}\sum_{k=1}^{T} r_k^4 \mathrm{e}^{j4\phi_k} = \frac{1}{T}\sum_{k=1}^{T}\big(z_1(k) + jz_2(k)\big)^4, \tag{3.29}$$

$$\hat{\gamma} = \frac{1}{T}\sum_{k=1}^{T} r_k^4 - 8 = \frac{1}{T}\sum_{k=1}^{T}\big(z_1^2(k) + z_2^2(k)\big)^2 - 8, \tag{3.30}$$

so that ξ_γ can be estimated from the above two:

$$\hat{\xi_\gamma} = \hat{\xi} \cdot \mathrm{sign}(\hat{\gamma}). \tag{3.31}$$

As a result, the following angle estimator arises:

$$\hat{\theta}_{\mathrm{EML}} = \frac{1}{4}\arg(\hat{\xi_\gamma}), \tag{3.32}$$

in which function "$\arg(\hat{\xi_\gamma})$" supplies the principal value of the argument of $\hat{\xi_\gamma}$ (i.e, the four-quadrant phase of $\hat{\xi_\gamma}$). The acronym *EML* stands for *Extended ML*. This name will find justification in the next section. Observe that no assumptions on the source pdf's have been made at all to arrive at this expression, which makes this estimator valid for *any* source distribution combination with *any* kurtosis value and *any* type of symmetry, as long as the sks is not zero. In the event of such a sum being zero, the magnitude of ξ would also be null and θ could not be estimated from it any more. More specifically, we will see in Section 3.3.5 when carrying out the asymptotic performance analysis that estimator (3.32) becomes inconsistent as $\gamma \to 0$.

Before closing this section, recall that a $\frac{\pi}{2}$-radian estimation range for the unknown parameter θ suffices (Section 3.2.1). Since $\arg(\cdot) \in \Omega_{2\pi} = [-\pi, \pi]$, the EML estimator (3.32) always yields angle estimates in the required range. Effectively, the argument of the term "$\xi \cdot \operatorname{sign}(\gamma)$" is 4θ. If $|\theta| > \frac{\pi}{4}$, then such argument will lie outside $\Omega_{2\pi}$. Hence, the "$\arg(\cdot)$" function would 'see' that angle but reduced to the principal-value interval $\Omega_{2\pi}$, that is, it would 'see' $4\hat{\theta} = 4\theta - 2\pi n \operatorname{sign}(\theta)$, for certain positive integer n such that $4\hat{\theta} \in \Omega_{2\pi}$. Consequently, the angle estimated by (3.32) would be $\hat{\theta} = \theta - \operatorname{sign}(\theta)\frac{n\pi}{2}$, which belongs to the equivalence class of valid solutions (3.6).

3.3.3 Connections

In this section we discuss the links between the EML estimator just presented and the works of Harroy-Lacoume, Comon and Bogner-Clarke.

Connection with the AML estimator

Let us relate expressions (3.19) and (3.32). The differences brought in by the arctangent and the argument functions in the respective estimation formulae are subtle but significant. Equation (3.19) can be written as:

$$\hat{\theta}_{\mathrm{AML}} = \frac{1}{4} \operatorname{arctg}\left(\frac{\mathbb{I}\mathrm{m}(\hat{\xi})}{\mathbb{R}\mathrm{e}(\hat{\xi})}\right), \tag{3.33}$$

where $\mathbb{R}\mathrm{e}(\cdot)$ and $\mathbb{I}\mathrm{m}(\cdot)$ represent, respectively, the real and imaginary part of their complex argument. First, let us assume that $\hat{\theta}_{\mathrm{AML}}$ is calculated without regard to the sign of the numerator and denominator of the arctangent argument, so that "$\operatorname{arctg}(\cdot)$" only produces two-quadrant phase values. Contrasting the latter expression with (3.32), it turns out that both provide the same estimates when $\angle\hat{\xi}_\gamma \in [-\pi/2, \pi/2]$, or, assuming that no error occurs in the estimation of ξ_γ, i.e., $\hat{\xi}_\gamma = \xi_\gamma$, and by virtue of (3.28), when $\theta \in [-\pi/8, \pi/8] = \Omega_{\pi/4}$. Estimator $\hat{\theta}_{\mathrm{AML}}$ is a particular case of $\hat{\theta}_{\mathrm{EML}}$ when $\theta \in \Omega_{\pi/4}$, and hence $\hat{\theta}_{\mathrm{AML}}$ is indeed applicable under the same source pdf conditions as $\hat{\theta}_{\mathrm{EML}}$. When $\theta \notin \Omega_{\pi/4}$ estimator (3.19) is biased. This was expected, since if the sign of the arctangent arguments is not taken into account, $\operatorname{arctg}(\cdot) \in [-\pi/2, \pi/2]$ and then (3.19) only provides estimates in the range $[-\pi/8, \pi/8]$. Effectively, the arctangent function in (3.33) obtains $4\theta - \pi$ when $4\theta > \pi/2$, and $4\theta + \pi$ when $4\theta < -\pi/2$. Consequently, if $\theta > \pi/8$, the expected bias is $-\pi/4$ radians, whereas if $\theta < -\pi/8$ such bias becomes $\pi/4$. For $\theta \notin \Omega_{\pi/4}$ one then needs to resort to expression (3.32).

In the second place, even if the sign of the arctangent arguments are considered, so that "arctg(·)" generates full four-quadrant phase values, estimator (3.33) only provides valid separation solutions when the sks is positive. Else, a bias arises in $\hat{\theta}_{\mathrm{AML}}$ relative to $\hat{\theta}_{\mathrm{EML}}$, which is $\pi/4$ radian when $\theta < 0$ and $-\pi/4$ radian when $\theta > 0$. In this respect, the effect of the sign function in (3.31)–(3.32) can alternatively be seen as applying an extra 45^{o} rotation to the estimate obtained for positive sks, so that the potential bias is avoided (this notion will be revisited when giving a geometrical interpretation of the method in Section 3.3.4). In essence, by virtue of the sign function the method is made transparent to the source pdf tailness. To emphasize the key role played by the sks, notice that such a factor also appears itself in Harroy-Lacoume's development, but vanishes when nulling the score function (see the term "$\mathrm{Lgc}_4(\rho)$" in [69, Section 3, page 171]).

In conclusion, it can be claimed that $\hat{\theta}_{\mathrm{EML}}$ extends $\hat{\theta}_{\mathrm{AML}}$ in the source pdf's which it is able to deal with. The constraints (symmetric sources with the same distribution and positive kurtosis in the range [0, 4]) under which the approximate ML estimator was originally inferred are alleviated in the proposed method, and reduced to a much weaker single condition on the sum of source kurtosis. Although $\hat{\theta}_{\mathrm{EML}}$ has not been derived from the ML principle, the term "ML" has got some significance. The condition for $\hat{\theta}$ to be the ML estimator of θ under the assumptions of [69] may be restated as $\mathrm{tg}(4\hat{\theta}) = \frac{\mathrm{Im}(\hat{\xi})}{\mathrm{Re}(\hat{\xi})}$, which is fulfilled by (3.32), and hence the latter is also the ML estimator of θ under the same assumptions. This claim will be mathematically proven in the next chapter.

First connection with Comon's formula

It is worth computing ξ as a function of the whitened-observation statistics. On the one hand, from (3.10) and (3.21):

$$\xi = \mathrm{E}[(z_1 + jz_2)^4] = (\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z) + j4(\kappa_{31}^z - \kappa_{13}^z), \tag{3.34}$$

and, on the other hand, from (3.10), (3.26) and (3.27):

$$\gamma = \mathrm{E}[(z_1^2 + z_2^2)^2] - 8 = \kappa_{40}^z + 2\kappa_{22}^z + \kappa_{04}^z. \tag{3.35}$$

Combining this expression with (3.24) yields:

$$\xi = \mathrm{e}^{j4\theta}\gamma = \mathrm{e}^{j4\theta}(\kappa_{40}^z + 2\kappa_{22}^z + \kappa_{04}^z). \tag{3.36}$$

The last equations state that θ can be determined from the 4th-order cumulants of the decorrelated measurements. In particular, the modulus of (3.34) and (3.36) must be equal, which leads to the following relationship among the 4th-order cumulants of the whitened sensor outputs:

$$(\kappa_{31}^{z} - \kappa_{13}^{z})^2 - \kappa_{22}^{z}(\kappa_{40}^{z} + \kappa_{04}^{z}) + 2(\kappa_{22}^{z})^2 = 0, \tag{3.37}$$

This relationship was originally deduced by Comon in [41] and [42] following different and more algebraic arguments. In the latter reference this relation is equivalently written as:

$$\rho^2 - \sigma + 2 = 0, \qquad \text{with} \begin{cases} \rho \triangleq \frac{\kappa_{31}^{z} - \kappa_{13}^{z}}{\kappa_{22}^{z}} \\ \sigma \triangleq \frac{\kappa_{40}^{z} + \kappa_{04}^{z}}{\kappa_{22}^{z}}. \end{cases} \tag{3.38}$$

Connection with Bogner-Clarke's estimator

In [39] a general class of estimators of parameter θ is unveiled. They are fundamentally based on geometric ideas inspired by the scatter diagram, ideas which are also applicable for the EML estimator (see Section 3.3.4). The general expression for the estimators proposed by Clarke resembles that of the EML (3.32):

$$\hat{\theta}_{\mathrm{CLA}} = \frac{1}{4} \arg\Big(\sum_k F(r_k) \mathrm{e}^{j4\phi_k}\Big), \tag{3.39}$$

where function $F(\cdot)$, so-called *radial weighting*, is a (possibly non-linear) function of the moduli r_k. The rationale behind this function is to improve the estimation robustness in the presence of outliers caused by impulsive noise [39, 118].

Though intuitive and insightful, the geometric notions upon which this estimator is based do not find in [39] solid statistical grounds. These geometric principles entirely rely on the empirical observation of scatter plots of symmetric sources and, as a consequence, can only be exploited when the sources have symmetric distributions. In contrast, by resorting to sound statistical foundations, we have been able to prove that the EML estimator is valid for any type of source symmetry. Also, remark that, according to the explanations of the previous sections, eqn. (3.39) is only applicable for positive source kurtosis sum (at least for $F(r) = r^4$, which corresponds to the radial weighting used in the EML), detail of which the author of [39] seems to be unaware.

In [118] this deficiency is finally realized, and partially surmounted by means of a "$\pm$" sign inside the argument of the angle function. However, it is still obscure how the sign uncertainty is resolved. We settle this matter by proving a) that such sign corresponds to the sign of source kurtosis sum, and b) that the latter, in turn, can be calculated from the available information as in (3.27), thus resulting in a fully-blind estimation procedure.

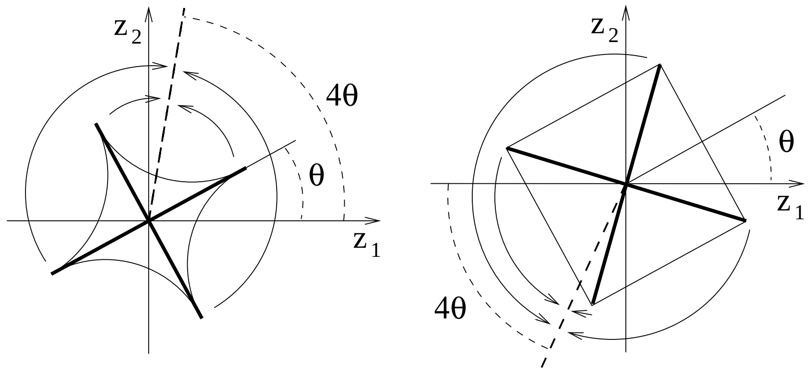
Finally, we want to stress that the original idea of exploiting the symmetries of the scatter diagram in the two-signal case dates back to Bogner [13], although no explicit estimation expression is actually provided therein, contrary to the claims in [39] and [118]. Also, in [118] estimator (3.39) is derived for $F(r) = r^4$ from the approximate ML function obtained by Harroy and Lacoume [eqn. (3.18)]. Yet the connection between the ML principle and the general family (3.39) is statistically unclear for arbitrary radial weighting $F(r)$, and stems only from intuitive geometrical reasons.

3.3.4 Geometrical interpretation

There exists a fairly enlightening geometrical rationale behind estimator expression (3.32) when the two sources are symmetrically distributed. Analogous geometric ideas were originally exposed in [13], and later in [39] and [118]. Firstly, consider a mixture of two leptokurtic pdf's, such as, for instance, two Laplacian distributions. Since the kurtosis of both sources is positive, it is guaranteed that the sum is. The jpdf of the unit-power sources exhibits highest values along the lines defined by the angles $\phi' = n\frac{\pi}{2}$, $n = 0, 1, 2, 3$. Accordingly, the source scatter diagram has a maximum point concentration along the same lines. On the other hand, as explained in Section 3.2.3, the jpdf of the whitened measurements is a θ-rad rotated version of that of the sources, θ being unknown. As a result, the new scatter diagram displays the highest density along $\phi = \theta + n\frac{\pi}{2}$, $n = 0, 1, 2, 3$. Now, assume that all the scatter diagram points are transformed according to

$$r_k e^{j\phi_k} \mapsto r_k^4 e^{j4\phi_k}. \tag{3.40}$$

In particular, all points previously cluttered around $\phi = \theta + n\frac{\pi}{2}$ will now be gathered together along $\phi = 4\theta$, as graphically depicted in Figure 3.2(a). In this way it can be



(a) (b)

Figure 3.2: EML centroid orientation (straight dashed line) for a mixture of two symmetric distributions: (a) both leptokurtic, (b) both platykurtic.

argued that a *centroid* calculated as the resultant mean point after the transformation, that is

$$\hat{\xi} = \frac{1}{T} \sum_k r_k^4 e^{j4\phi_k}, \tag{3.41}$$

will show that orientation. Therefore, expression (3.32) will provide the required angle with which the whitened jpdf is rotated with respect to the true source jpdf.

In the second place, let us now consider two platykurtic distributions playing the role of sources. Hence, their kurtosis sum is negative. For simplicity, assume two sinusoidal signals. In such case, the maximum concentration in the source scatter diagram occurs along the lines $\phi' = \frac{\pi}{4} + n\frac{\pi}{2},\ n = 0, 1, 2, 3$. Therefore, in the scatter diagram of the signals after decorrelation this high density will be shown along $\phi = \theta + \frac{\pi}{4} + n\frac{\pi}{2}$, $n = 0, 1, 2, 3$. These points will all clutter round $\phi = 4\theta + \pi$ when transformed as indicated by (3.40), and so the centroid (3.41) will also have this orientation [Figure 3.2(b)]. In conclusion, the centroid projected across the origin, i.e., $-\xi$, forms an angle 4θ with respect to the x-axis. That is the purpose of the "sign(·)" term in the estimator expression (3.31)–(3.32) when $\kappa_{40}^x + \kappa_{04}^x < 0$: to rotate the sample centroid $\hat{\xi}$ (3.29) by π radians so that its argument becomes an estimate of 4θ.

For combinations of platykurtic and leptokurtic distributions, the geometrical interpretation becomes less intuitive, but the results obtained in Section 3.3.2 indicate that those cases can also be explained in terms of the above two instances, comprising sources with positive and negative kurtosis sum. However, a degradation in the EML-estimator performance is expected as the source kurtosis become closer in absolute value but opposite in sign, as analyzed in the next section.

3.3.5 Statistical characterization

Having introduced the expression of the EML estimator, highlighted its links to similar methods, and examined its behaviour from a geometrical point of view, a study of its statistical properties is in order. The present section is devoted to analyzing its asymptotic performance, that is, its estimation characteristics when a large number of samples is processed.[2] Specifically, our main objective is to obtain an analytical expression for the estimator asymptotic pdf.

Bias

Before entering the asymptotic domain, observe first that the EML estimator (3.32) is *unbiased for any sample size* when at least one of the sources has a symmetric distribution. The proof of this claim flows along the same lines as that offered in [69, Section 4.1] for the approximate ML estimator (3.19). If there is an asymmetric source, let its corresponding axis form an angle θ with respect to the whitened observations frame of reference (the reasoning is totally analogous if a different axis configuration is chosen). Consider a set of observations (r_k, ϕ_k). From this set, parameters $\hat{\xi}^{(1)}$ and $\hat{\gamma}^{(1)}$ are computed by means of (3.29) and (3.30), respectively, yielding an angle estimator $\hat{\theta}^{(1)}_{\mathrm{EML}}$. Now, consider the set of symmetric (with respect to the true angle θ) observations $(r_k, 2\theta - \phi_k)$. The new parameters are related to the previous ones via $\hat{\gamma}^{(2)} = \hat{\gamma}^{(1)}$ and $\hat{\xi}^{(2)} = \hat{\xi}^{(1)*} \mathrm{e}^{j8\theta}$, symbol * denoting complex conjugation. Then, from (3.31)–(3.32), $\hat{\theta}^{(2)}_{\mathrm{EML}} = 2\theta - \hat{\theta}^{(1)}_{\mathrm{EML}}$.[3] By symmetry of the source(s), $\mathrm{E}[\hat{\theta}^{(1)}_{\mathrm{EML}}] = \mathrm{E}[\hat{\theta}^{(2)}_{\mathrm{EML}}]$. It follows that $\mathrm{E}[\hat{\theta}_{\mathrm{EML}}] = \theta$, for any sample size. Remark that for this result to hold we do

[2]The usefulness of asymptotic performance analysis is pointed out in [119].
[3]Plus a potential irrelevant $\pm\pi/2$ bias term.

not need to assume that both sources possess equal symmetric distribution as in [69], but just that one source be symmetric.

Basic asymptotic results

If the sources are both asymmetric, the above result is not generally true. However, under basic regularity conditions other interesting conclusions about estimator (3.32) can be drawn, which hold for all kinds of source pdf. Effectively, assuming independent and identically distributed (i.i.d.) observations, the strong law of large numbers (SLLN) [113, 120] ensures that

$$\hat{\xi} \xrightarrow[T\to\infty]{\text{wp1}} \xi \tag{3.42}$$

and

$$\hat{\gamma} \xrightarrow[T\to\infty]{\text{wp1}} \gamma, \tag{3.43}$$

and in compliance with equations (3.28)–(3.32),

$$\hat{\theta}_{\text{EML}} \xrightarrow[T\to\infty]{\text{wp1}} \theta, \tag{3.44}$$

where $\xrightarrow{\text{wp1}}$ denotes convergence with probability one. Hence, the estimator is *strongly consistent* and *asymptotically unbiased*, provided $\gamma \neq 0$.

Asymptotic pdf

In order to arrive at a more complete characterization of the estimator, consider first that, from relationship (3.13), centroid $\hat{\xi}$ can be written as:

$$\hat{\xi} = \frac{1}{T}\sum_{k=1}^{T} r_k^4 \mathrm{e}^{j4\phi_k} = \left\{\frac{1}{T}\sum_{k=1}^{T} r_k^4 \mathrm{e}^{j4\phi'_k}\right\} \mathrm{e}^{j4\theta} = \hat{\xi}_x \mathrm{e}^{j4\theta}, \tag{3.45}$$

in which

$$\hat{\xi}_x = \frac{1}{T}\sum_{k=1}^{T} r_k^4 \mathrm{e}^{j4\phi'_k} = \frac{1}{T}\sum_{k=1}^{T}\big(x_1(k) + jx_2(k)\big)^4. \tag{3.46}$$

Then the estimator becomes

$$\hat{\theta} = \frac{1}{4}\arg\big(\hat{\xi}\cdot \text{sign}(\hat{\gamma})\big) = \theta + \frac{1}{4}\arg\big(\hat{\xi}_x \cdot \text{sign}(\hat{\gamma})\big) \tag{3.47}$$

so that

$$\Delta\hat{\theta} = \hat{\theta} - \theta = \frac{1}{4}\arg\big(\hat{\xi}_x \cdot \operatorname{sign}(\hat{\gamma})\big), \tag{3.48}$$

That is, the estimation error, $\Delta\hat{\theta} = \hat{\theta} - \theta$, depends exclusively on the source signals (note that $\hat{\gamma}$ only depends on them too) and is independent of the specific value of the rotation angle θ. Our objective, then, is reduced to obtaining the pdf of the error $\Delta\hat{\theta}$.

In the second place, let us assume that the sign of the sks is properly estimated, which is a reasonable assumption for a high enough sample size T and/or when $|\gamma| = |\kappa_{40}^x + \kappa_{04}^x|$ is not too small. Let us also assume that this sign is positive. Calling

$$\delta = 4\Delta\hat{\theta} = \arg(\hat{\xi}_x), \tag{3.49}$$

we have

$$p_{\Delta\hat{\theta}}(\Delta\hat{\theta}) = 4p_\delta(4\Delta\hat{\theta}), \qquad |\delta| \leqslant \pi, \quad |\Delta\hat{\theta}| \leqslant \frac{\pi}{4}, \tag{3.50}$$

so that it is sufficient to obtain $p_\delta(\delta)$. If the sks is negative, the actual pdf will be a π-radian circularly shifted version of $p_\delta(\delta)$ above, due to the sign function in (3.48). At this point, it is convenient to decompose $\hat{\xi}_x$ in its real and imaginary parts:

$$\hat{\xi}_x = \omega_1 + j\omega_2, \tag{3.51}$$

where ω_1 and ω_2 are obtained by developing the fourth power in (3.46):

$$\omega_1 = \frac{1}{T}\sum_{k=1}^{T}\big(x_1^4(k) - 6x_1^2(k)x_2^2(k) + x_2^4(k)\big) = \frac{1}{T}\sum_{k=1}^{T}\tilde{\omega}_1(k), \tag{3.52}$$

and

$$\omega_2 = \frac{1}{T}\sum_{k=1}^{T}4\big(x_1^3(k)x_2(k) - x_1(k)x_2^3(k)\big) = \frac{1}{T}\sum_{k=1}^{T}\tilde{\omega}_2(k). \tag{3.53}$$

With the help of (3.10) and [120], the statistical properties of $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are obtained as follows. In the first place, the mean and variance of $\tilde{\omega}_1$ are given by:

$$\mathrm{E}[\tilde{\omega}_1] = \mu_{40}^x + \mu_{04}^x - 6 = \kappa_{40}^x + \kappa_{04}^x = \gamma, \tag{3.54}$$

$$\begin{aligned}
\tilde{\sigma}_1^2 &\triangleq \mathrm{Var}[\tilde{\omega}_1] = \\
&= \big(\mu_{80}^x - (\mu_{40}^x)^2\big) - 12(\mu_{60}^x - \mu_{40}^x) + 36(\mu_{40}^x\mu_{04}^x - 1) - 12(\mu_{06}^x - \mu_{04}^x) + \big(\mu_{08}^x - (\mu_{04}^x)^2\big) = \\
&= (\kappa_{80}^x + \kappa_{08}^x) + 16(\kappa_{60}^x + \kappa_{06}^x) + 56(\kappa_{50}^x\kappa_{30}^x + \kappa_{05}^x\kappa_{03}^x) + 34\big((\kappa_{40}^x)^2 + (\kappa_{04}^x)^2\big) + \\
&\quad + 160\big((\kappa_{30}^x)^2 + (\kappa_{03}^x)^2\big) + 144(\kappa_{40}^x + \kappa_{04}^x) + 36\kappa_{40}^x\kappa_{04}^x + 192.
\end{aligned} \tag{3.55}$$

Analogously, the first and second order statistical parameters of $\tilde{\omega}_2$ are:

$$\mathrm{E}[\tilde{\omega}_2] = 0, \tag{3.56}$$

$$\begin{aligned}\tilde{\sigma}_2^2 \triangleq \mathrm{Var}[\tilde{\omega}_2] &= 16(\mu_{60}^x - 2\mu_{40}^x\mu_{04}^x + \mu_{06}^x) = \\ &= 16\big[(\kappa_{60}^x + \kappa_{06}^x) + 9(\kappa_{40}^x + \kappa_{04}^x) + 10\big((\kappa_{30}^x)^2 + (\kappa_{03}^x)^2\big) - 2\kappa_{40}^x\kappa_{04}^x + 12\big].\end{aligned} \tag{3.57}$$

For two Gaussian sources, these statistical parameters become:

$$\gamma = 0, \qquad \tilde{\sigma}_1^2 = \tilde{\sigma}_2^2 = 192. \tag{3.58}$$

The covariance of $\tilde{\omega}_1(k)$ and $\tilde{\omega}_2(k)$ is

$$\tilde{c}_{12} \triangleq \mathrm{E}[\tilde{\omega}_1\tilde{\omega}_2] - \mathrm{E}[\tilde{\omega}_1]\mathrm{E}[\tilde{\omega}_2] = -28(\mu_{50}^x\mu_{03}^x - \mu_{30}^x\mu_{05}^x) = -28(\kappa_{50}^x\kappa_{03}^x - \kappa_{30}^x\kappa_{05}^x), \tag{3.59}$$

and so these variables are not necessarily uncorrelated. However, the sole existence of a symmetric source makes $\tilde{c}_{12}$ vanish, and the same applies when both sources have identical distribution, for example.

Next, from (3.52) and (3.53), and assuming that each source signal is temporally white [property which is inherited by statistics $\tilde{\omega}_1(k)$ and $\tilde{\omega}_2(k)$], the central limit theorem (CLT) [113, 120] states that, for a high enough sample size T, ω_1 can be approximated by a Gaussian distribution of mean γ and variance $\sigma_1^2 = \tilde{\sigma}_1^2/T$, and ω_2 by another Gaussian variable of zero mean and variance $\sigma_2^2 = \tilde{\sigma}_2^2/T$, i.e.,

$$\begin{aligned}\omega_1 &\xrightarrow[T\to\infty]{\mathrm{d}} \mathrm{N}(\gamma,\, \sigma_1^2) = \mathrm{N}(\gamma,\, \tilde{\sigma}_1^2/T), \\ \omega_2 &\xrightarrow[T\to\infty]{\mathrm{d}} \mathrm{N}(0,\, \sigma_2^2) = \mathrm{N}(0,\, \tilde{\sigma}_2^2/T),\end{aligned} \tag{3.60}$$

where $\xrightarrow{\mathrm{d}}$ indicates convergence in distribution. On the other hand, for temporally white sources, $\mathrm{E}[\tilde{\omega}_1(k)\,\tilde{\omega}_2(n)] = \mathrm{E}[\tilde{\omega}_1(k)]\mathrm{E}[\tilde{\omega}_2(n)] = 0$ for all $k \neq n$, and thus the covariance of ω_1 and ω_2 is given by

$$c_{12} \triangleq \mathrm{E}[\omega_1\omega_2] - \mathrm{E}[\omega_1]\mathrm{E}[\omega_2] = \frac{\tilde{c}_{12}}{T}. \tag{3.61}$$

If $\tilde{c}_{12} = 0$, then ω_1 and ω_2 are uncorrelated. But even if the coefficient $\tilde{c}_{12}$ is far from zero, c_{12} cancels as T tends to infinity, and hence ω_1 and ω_2 are at least asymptotically

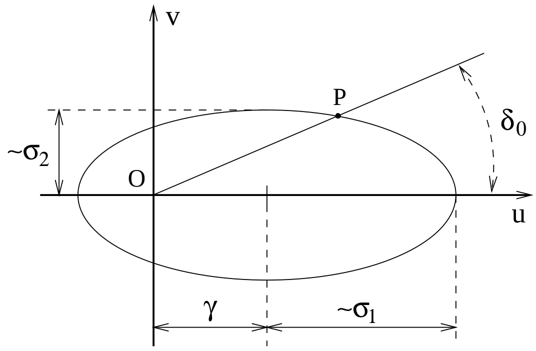


Figure 3.3: Contour lines of the asymptotic jpdf of ω_1 and ω_2, the source centroid coordinates in the complex plane.

uncorrelated in any case. In conclusion, the asymptotic jpdf of the real and imaginary parts of $\hat{\xi}_\gamma$ is given by the product of their marginal pdf's:

$$p_{\omega_1\omega_2}(u,\, v) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{(u-\gamma)^2}{2\sigma_1^2} - \frac{v^2}{2\sigma_2^2} \right\}. \tag{3.62}$$

The contour lines of this pdf are ellipses of semi-axes proportional to σ_1 and σ_2 centred at $(\gamma,\, 0)$, as depicted in Figure 3.3.

Having obtained the jpdf of ω_1 and ω_2, relationships (3.49) and (3.51) suggest to try the variable change:

$$(u,\, v) \mapsto (r,\, \delta) \quad : \quad \begin{cases} u = r\cos\delta \\ v = r\sin\delta \end{cases} \tag{3.63}$$

which yields for the marginal pdf of δ:

$$p_\delta(\delta) = \int_0^\infty \frac{r}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{(r\cos\delta-\gamma)^2}{2\sigma_1^2} - \frac{(r\sin\delta)^2}{2\sigma_2^2} \right\} dr. \tag{3.64}$$

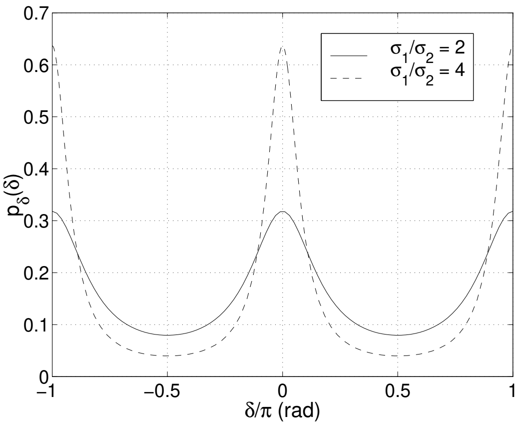
Although this expression admits an analytical solution, as will be seen later on, let us first derive further simplifications to gain a deeper insight into its behaviour as a function of parameters γ, σ_1 and σ_2. Fig. 3.3 can still be viewed as an approximation to a contour line of the integrand in (3.64). Roughly speaking, the value of p_δ at a particular angle δ_0 corresponds to the integral with respect to r of a slice of the function (3.62) along the angle $\delta = \delta_0$. Therefore, one could think of this value as being related to the distance from the origin to the point on the contour line located at $\delta = \delta_0$, i.e., distance OP in Fig. 3.3. As a result, it is easy to see now that for $\sigma_1 > \sigma_2$ and $\gamma > 0$ the distribution of δ will concentrate more closely around its mode $\delta = 0$ as $\frac{\gamma}{\sigma_1}$ increases. The same outcome is obtained if the ratio $\frac{\sigma_1}{\sigma_2}$ is increased. On the other hand, if $\sigma_1 < \sigma_2$, the distribution of δ will show two modes, neither at $\delta = 0$, though symmetric with respect to that point. Analogous results are obtained for $\gamma < 0$, but in that case the pdf of δ will exhibit one mode at $\delta = \pm\pi$ for $\sigma_1 > \sigma_2$ (this is the statistical counterpart of the geometric justification for the term "sign$(\cdot)$" given in Section 3.3.4), or two modes symmetrically located for $\sigma_1 < \sigma_2$. For $\sigma_1 < \sigma_2$ the two modes merge into only one as $\frac{|\gamma|}{\sigma_1}$ increases. As an example to illustrate this intuitive reasoning, consider two Gaussian sources, for which, according to (3.58), $\gamma = 0$ and $\sigma_1 = \sigma_2$, thus the contour curves become circumferences centred at the origin. This gives a $p_\delta(\delta)$ uniformly distributed between $-\pi$ and π.

The main result derived from these geometrical considerations is that the distribution of δ (and hence of $\Delta\hat{\theta}$) becomes more 'peaky' around 0 or $\pm\pi$ (depending on whether γ is positive or negative, respectively) as the ratios $\frac{|\gamma|}{\sigma_1}$ and $\frac{\sigma_1}{\sigma_2}$ become larger. In order to give a first mathematical support to this claim, one can readily work out the analytical expression of p_δ when $\gamma = 0$:

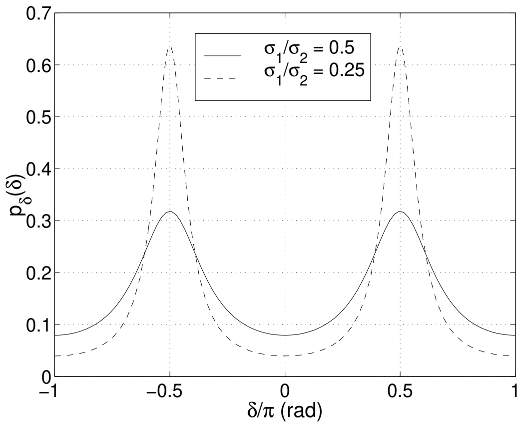
$$p_\delta(\delta) = \frac{\frac{\sigma_1}{\sigma_2}(1 + \mathrm{tg}^2\delta)}{2\pi\left[1 + (\frac{\sigma_1}{\sigma_2}\mathrm{tg}\,\delta)^2\right]}. \tag{3.65}$$

Then, for $\sigma_1 > \sigma_2$, the maximum value of p_δ is $\frac{\sigma_1}{\sigma_2}/(2\pi)$ and occurs at $\delta = 0, \pm\pi$, whereas the minimum is $1/(2\pi\frac{\sigma_1}{\sigma_2})$ at $\delta = \pm\pi/2$. The ratio of the maximum to the minimum value is thus $(\frac{\sigma_1}{\sigma_2})^2$, confirming that the distribution becomes more 'peaky' as $\frac{\sigma_1}{\sigma_2}$ increases, as seen in Figure 3.4(a). If, on the contrary, $\sigma_1 < \sigma_2$, the maximum value becomes $1/(2\pi\frac{\sigma_1}{\sigma_2})$ at $\delta = \pm\pi/2$ and the minimum $\frac{\sigma_1}{\sigma_2}/(2\pi)$ at $\delta = 0, \pm\pi$, as shown in Figure 3.4(b). The pdf is flat if and only if $\sigma_1 = \sigma_2$, as occurs, for instance, for Gaussian sources [see expression (3.58)].

All this preceding analysis establishes the behaviour of p_δ as a function of $\frac{\gamma}{\sigma_1}$ and $\frac{\sigma_1}{\sigma_2}$ approximately. Now, it will be seen how these trends are actually exhibited by the analytical expression of the pdf. With the help of the symbolic mathematics package MAPLE™ [36], and after some tedious algebraic manipulations, one can arrive from



(a)



(b)

Figure 3.4: Pdf of δ for $\gamma = 0$ [eqn. (3.65)]: (a) $\sigma_1/\sigma_2 > 1$, (b) $\sigma_1/\sigma_2 < 1$.

(3.64) at the following general analytical expression for the pdf of δ:

$$\begin{aligned} p_\delta(\delta) = & \frac{\frac{\sigma_1}{\sigma_2}(1+\mathrm{tg}^2\delta)}{2\pi\left[1+(\frac{\sigma_1}{\sigma_2}\mathrm{tg}\,\delta)^2\right]}\exp\left(\frac{-\gamma^2}{2\sigma_1^2}\right) + \\ & + \frac{\gamma(1+\mathrm{tg}^2\delta)\,\mathrm{sign}(\cos\delta)}{\sqrt{8\pi}\,\sigma_2\left[1+(\frac{\sigma_1}{\sigma_2}\mathrm{tg}\,\delta)^2\right]^{3/2}}\exp\left(\frac{-\gamma^2\mathrm{tg}^2\delta}{2\sigma_2^2\left[1+(\frac{\sigma_1}{\sigma_2}\mathrm{tg}\,\delta)^2\right]}\right)\cdot \\ & \cdot\left\{1+\mathrm{erf}\left(\frac{\gamma\,\mathrm{sign}(\cos\delta)}{\sqrt{2}\,\sigma_1\sqrt{1+(\frac{\sigma_1}{\sigma_2}\mathrm{tg}\,\delta)^2}}\right)\right\}. \end{aligned} \tag{3.66}$$

Observe that for $\gamma = 0$ the above formula simplifies into (3.65). The asymptotic results derived earlier may also be deduced from (3.66). Effectively, this pdf is in all cases symmetric with respect to $\delta = 0$, and hence $\mathrm{E}[\delta] = 0$, that is, the EML estimator is unbiased for large enough sample size. Also, as T increases, both σ_1 and σ_2 tend to zero, and so the distribution of ω_1 and ω_2 tends to a point of probability mass 1 located at $(\gamma, 0)$ in the (ω_1, ω_2)-plane. In consequence, the EML estimator shows strong consistency, as long as $\gamma \neq 0$, i.e., $\kappa_{40}^x + \kappa_{04}^x \neq 0$. For $\gamma = 0$, since $\frac{\sigma_1}{\sigma_2} = \frac{\tilde\sigma_1}{\tilde\sigma_2}$, which is a constant for a given pair of source distributions, according to (3.65) the estimator pdf does not vary as the sample size increases. Thereby the estimator loses its consistency property when the sks is null.

Finally, in order to circularly shift the pdf expression by π radians, which is necessary when $\kappa_{40}^x + \kappa_{04}^x < 0$ as discussed at the beginning of this section, it suffices to use $|\gamma|$ instead of γ in (3.66).

Gaussian approximation

In the i.i.d. case, the variances σ_1^2 and σ_2^2 decrease as T increases, while γ remains fixed (assumed finite). For T large enough, $|\gamma| \gg \sigma_1, \sigma_2$, and then the following approximations hold. In the first place, $\exp\big(-\gamma^2/(2\sigma_1^2)\big) \approx 0$. Also, the $p_\delta(\delta)$ concentrates around $\delta \approx 0$, so that $\operatorname{tg}\delta \approx \delta$, $\operatorname{sign}(\cos\delta) \approx 1$, $(1 + \operatorname{tg}^2\delta) \approx 1$ and $[1 + (\sigma_1/\sigma_2 \operatorname{tg}\delta)^2] \approx 1$. Hence, the error function in (3.66) tends to one, whereas the second exponential can be approximated by $\exp\big(-\gamma^2\delta^2/(2\sigma_2^2)\big)$. As a consequence of these simplifications, asymptotic pdf (3.66) reduces to:

$$p_\delta(\delta) \underset{T\to\infty}{\approx} \hat{p}_\delta(\delta) = \frac{1}{\sqrt{2\pi}\,\sigma_2/|\gamma|} \exp\left\{\frac{-\delta^2}{2(\sigma_2/\gamma)^2}\right\}. \tag{3.67}$$

That is:

$$\delta \xrightarrow[T\to\infty]{\mathrm{d}} \mathrm{N}(0,\, \sigma_2^2/\gamma^2), \tag{3.68}$$

From this Gaussian approximation the asymptotic variance of the estimators can be derived at once, which was not easy to obtain from pdf (3.66) directly. This variance reads:

$$\operatorname{Var}[\hat{\theta}_{\mathrm{EML}}] \xrightarrow[T\to\infty]{} \sigma_{\mathrm{EML}}^2 = \frac{\mu_{60}^x + \mu_{06}^x - 2\mu_{40}^x\mu_{04}^x}{T(\kappa_{40}^x + \kappa_{04}^x)^2}. \tag{3.69}$$

The quality of this Gaussian approximation is visualized in Figure 3.5. It represents the contour lines of the normalized square difference $\varepsilon^2 = \int(\hat{p}_\delta - p_\delta)^2 d\delta / \int p_\delta^2\, d\delta$, for a range of values of γ, σ_1 and σ_2. It was obtained by trapezoidal numerical integration, with 1000 equispaced points between $\max(-\pi, -3\sigma)$ and $\min(\pi, 3\sigma)$, σ^2 being the variance of pdf (3.67), i.e., σ_2^2/γ^2. Approximately, $\varepsilon^2 < 1\%$ for $\sigma_1 > 0.6\,\sigma_2$ and $|\gamma| > 2.5\,\sigma_1$. Since (asymptotically) $\sigma_1/\sigma_2 = \tilde{\sigma}_1/\tilde{\sigma}_2$ and $\gamma/\sigma_1 = T\gamma/\tilde{\sigma}_1$, the information derived from Fig. 3.5 is useful in evaluating the sample size required to achieve a certain approximation quality, for a given pair of source distributions.

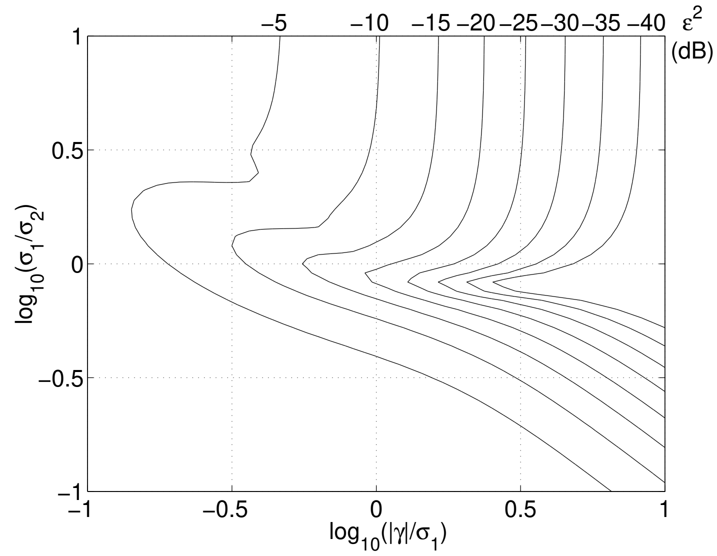


Figure 3.5: Accuracy of the Gaussian approximation $\hat{p}_\delta$ (3.67) to the analytic asymptotic pdf p_δ (3.66) of the EML estimator. The plot represents the contour lines of $\varepsilon^2_{\mathrm{dB}} = 10\log_{10}\left(\int(\hat{p}_\delta - p_\delta)^2 d\delta / \int p_\delta^2\, d\delta\right)$.

Orthogonal invariance

We now prove that the EML estimator is orthogonal invariant for any sample size. The key point is that the sampling error is independent of the actual value of θ, as conveyed by error equation (3.48).

Let matrix Z_T be a realization of the whitened observations composed of a block of T samples. Recall (Section 2.3) that an estimator $\widehat{\mathbb{Q}}$ used to find rotation Q is said to satisfy the *orthogonal invariance* property, if

$$\forall U \text{ unitary}, \quad \widehat{\mathbb{Q}}(UZ_T) = U\widehat{\mathbb{Q}}(Z_T)V, \tag{3.70}$$

for almost any realization of Z_T and for some irrelevant quasiidentity matrix V [24]. The application on Z_T of estimator (3.32) yields: $\widehat{Q} = \widehat{\mathbb{Q}}(Z_T) = QLF$, where matrix L represents the estimation error (3.48), which is independent of Q, and F a quasi-identity matrix as in (3.8), modelling the potential alternative solutions (3.6). Now, pre-multiplying the data Z_T by U is tantamount to pre-multiplying the original rotation Q by the same matrix. But unitary matrix U can be regarded as an additional rotation of angle $\Delta\theta$, and hence matrix UQ is just a rotation of angle $\theta + \Delta\theta$. Estimator (3.32) would recover a rotation of that angle sum plus the error $\Delta\hat{\theta}$, independent of $\theta + \Delta\theta$ and thus represented again by matrix L, possibly post-multiplied by another quasiidentity matrix F'. That is, the estimated transformation would be $\widehat{Q}' = UQLF' = U\widehat{Q}V$, with $V = F^{-1}F'$. Since the product of two quasiidentity matrices is another quasiidentity, and so is the inverse of such a matrix, from definition (3.70) it follows that (3.32) is orthogonal invariant.

Further, in the absence of noise, a separation algorithm based on (3.32) would show the uniform performance property [24] (Section 2.3), whereby the reconstructed sources do not depend on the mixing matrix M, but only on the particular realization of the source signals.

3.3.6 Computational complexity

Finally, we are interested in evaluating how costly EML estimator (3.32) is computationally, i.e., the number of flops it takes to produce a result. The computational load stems, mainly, from the calculation of the sample centroid $\hat{\xi}$ and the sample source kurtosis sum $\hat{\gamma}$. Two cases must be distinguished, depending on whether these values are estimated from the scatter-plot expressions (3.29)–(3.30) or from the cumulants, as in eqns. (3.34) and (3.36). In any case, we consider (Section 2.4.5) that a complex product takes 4 flops, since as many real products are involved. Signal blocks are assumed to be composed of T samples.

1. Scatter-plot based expressions. Calculation of $\hat{\xi}$: term $(z_1 + jz_2)^4$: 2 complex products = 8 flops per sample. Calculation of $\hat{\gamma}$: term $(z_1^2 + z_2^2)^2$: no additional products, by taking advantage of partial results in the computation of the previous term. Overall: around $8T$ flops.

2. Cumulants. The sample estimates of expressions (3.10) are regarded:

$$\hat{\kappa}_{40}^z = \sum_k z_1(k)^4 - 3, \quad \hat{\kappa}_{31}^z = \sum_k z_1(k)^3 z_2(k), \quad \hat{\kappa}_{22}^z = \sum_k z_1(k)^2 z_2(k)^2 - 1$$
$$\hat{\kappa}_{04}^z = \sum_k z_2(k)^4 - 3, \quad \hat{\kappa}_{13}^z = \sum_k z_1(k) z_2(k)^3 \tag{3.71}$$

   For each sample, the terms in the sums can be calculated in optimized form as follows, each step taking 1 flop (see also [48, Section 4.3]): $z_{11} = z_1^2$; $z_{22} = z_2^2$;

$z_{12} = z_1 z_2$; $z_{1111} = z_{11}^2$; $z_{1112} = z_{11} z_{12}$; $z_{1122} = z_{11} z_{22}$; $z_{1222} = z_{12} z_{22}$; and $z_{2222} = z_{22}^2$. So the overall computation of the cumulant and also, roughly, of $\hat{\xi}$ and $\hat{\gamma}$ by this procedure is about $8T$ flops.

Therefore, the scatter-plot based centroid form and the cumulant form are just as expensive computationally, but the former presents the advantage of its appealing notational compactness.

3.3.7 Simulations

The purpose of this section is to illustrate the properties and performance of the suggested estimator, as well as to validate the theoretical asymptotic analysis, through a variety of simulations. The effects of a whitening process prior to the application of the estimator are also examined.

Orthogonal invariance

First, it is demonstrated that the performance of the EML estimator is independent of the particular value of the true angle θ. A pair of distributions are chosen for the source signals. At each angle value, separations are performed on 1000-sample mixtures over 100 independent Monte Carlo (MC) runs, with a fixed orthogonal mixing matrix (the rotation of the current angle). Following each separation, a performance index is computed from the results, and then averaged after the 100 realizations. Care must be taken when choosing the performance index. One could think of evaluating the mean squared error (MSE) of the estimated angle, $\mathrm{MSE}_{\hat{\theta}} \triangleq \mathrm{E}[(\hat{\theta} - \theta)^2]$. However, due to the equivalence class of valid solutions (3.6) and the fact that $\hat{\theta}_{\mathrm{EML}} \in \Omega_{\pi/2}$, the variance and MSE of $\hat{\theta}$ increases when θ is near or outside the limits of the $\frac{\pi}{2}$-rad estimation interval, even though satisfactory separations may still be attained. Instead, we choose the ISR, which, as commented in Section 2.4.3, is a measure of the average interference caused by the unwanted sources in the recovery of the desired sources. In the two-signal case, and since $\hat{\mathbf{x}} = \mathbf{s} = \widehat{Q}^{\mathrm{T}} Q \mathbf{x} = \mathcal{Q}(\theta - \hat{\theta})\mathbf{x}$ and the true sources are assumed unit-variance, it is easy to see that the ISR becomes a good approximation of the MSE when $\hat{\theta} \approx \theta + n\pi/2$, $n \in \mathbb{Z}$, i.e., when valid separations solutions are obtained. For example, assuming that $\hat{\theta} \approx \theta$, the recovered sources are

$$\hat{\mathbf{x}} = \begin{bmatrix} \cos(\hat{\theta} - \theta) & -\sin(\hat{\theta} - \theta) \\ \sin(\hat{\theta} - \theta) & \cos(\hat{\theta} - \theta) \end{bmatrix} \mathbf{x} \approx \begin{bmatrix} 1 & -(\hat{\theta} - \theta) \\ \hat{\theta} - \theta & 1 \end{bmatrix} \mathbf{x}, \tag{3.72}$$

and hence ISR $\approx \mathrm{E}[(\hat{\theta} - \theta)^2]$. The mean squared value of ISR in dB is defined as $\mathrm{ISR}^2(\mathrm{dB}) = 5\log_{10}(\mathrm{E}[\mathrm{ISR}^2])$.[4] Therefore, by using this measure both location (bias) and dispersion (variance) information about ISR are condensed in the results. The results are shown in Figure 3.6, for two source combinations: Gaussian–Rayleigh and exponential–uniform. The performance is constant with θ. In addition, the asymptotic variance of $\hat{\theta}$ found in (3.69) (and worked out for each source combination with the help of Tables A.1 and A.2 in Appendix A.4) is an accurate prediction of the empirical ISR.[5]

Performance for several source distribution pairs

Next, several MC simulations are run to test and compare the new estimator performance. Three different combinations of source pdf's as regards their tail or kurtosis sign are considered: both short, long-short and both long, together with three different symmetry combinations: both symmetric, symmetric-asymmetric and both asymmetric. That makes a total of nine distinct distribution pairs for the source signals. The actual source pdf's employed are (in parentheses the abbreviations used in the results table): uniform ("Uni"), exponential ("Exp"), Laplacian ("Lap"), Rayleigh ("Ray") and a skewed triangular distribution ("Tri"). The statistical parameters of these distribution are summarized in Appendix A.4.

For each distribution pair, zero-mean unit-variance source signal realizations are created from 5000 i.i.d. samples, and possible remains of statistical dependence up to 4th-order are removed by using the ICA procedure developed in [48] (the HOEVD

[4] The justification for the "5" factor lies in the fact that ISR is already a quadratic quantity [eqns. (2.63)–(2.65)].

[5] The fitness is even more precise if the empirical variance or MSE of $\hat{\theta}$ is plotted instead of ISR, although for the reasons given above this is not always meaningful. Furthermore, note that $10\log_{10}(\mathrm{E}[\mathrm{ISR}])$ will generally be more faithful to the theoretical variance or MSE predictions than $5\log_{10}(\mathrm{E}[\mathrm{ISR}^2])$, since this latter index also reflects how disperse the realizations of ISR (and, by approximation, of MSE) appear after the MC iterations. These comments should be borne in mind hereafter every time that ISR results are presented and contrasted with analytic angle variances or MSE's.

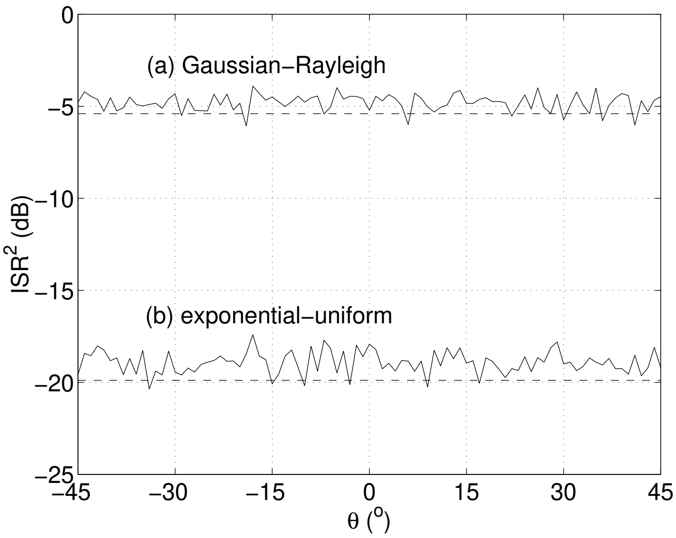


Figure 3.6: Performance of the EML estimator as a function of the true angle θ, $T =$ 1000 samples/signal. Solid lines: ISR^2 from the MC simulations, 100 MC runs per angle value. Dashed lines: variance of $\hat{\theta}$ predicted from (3.69) ($10\log_{10}\sigma^2_{\mathrm{EML}}$). Source distributions: (a) Gaussian–Rayleigh , (b) exponential–uniform.

method, see Section 2.5.4). Then a rotation matrix of fixed angle, $\theta = 15^{\mathrm{o}}$, is applied to the set of source signals, giving a hypothetical set of whitened sensor outputs. The estimation of the rotation angle is carried out through expressions (3.31)–(3.32), where the sample estimates $\hat{\xi}$ and $\hat{\gamma}$ are obtained from the whitened sensor data by means of (3.29) and (3.30). From each realization, two parameters are computed: the error $(\hat{\theta}-\theta)$ expressed in degrees, and the values of $\hat{\gamma}$ in order to check how they approximate the sks $(\kappa^x_{40} + \kappa^x_{04})$. The mean, standard deviation, maximum and minimum value for these two parameters computed over 100 MC runs are summarized in Table 3.1.

In light of these results, the new estimator can be judged as successful in estimating the unknown rotation angle θ. This is manifested in the low bias and variance of $\hat{\theta}$. Also, $\hat{\gamma}$ approximates the sks very accurately in all cases (e.g., for uniform–exponential sources is approximately equal to $-1.2 + 6 = 4.8$).

It is interesting to compare the variance obtained here with the bound given by (3.20). For example, for two Laplacian distributions ($\kappa^x_{40} = \kappa^x_{04} = 3$) and $T = 5000$ samples, equation (3.20) predicts a lower bound for the standard deviation of 0.468^{o}. However, from Table 3.1 the empirical standard deviation obtained for this case turns

DISTR.	BOTH SYMMETRIC			SYMMET. – ASYMMET.			BOTH ASYMMETRIC		
TYPE	PAIR	$\hat{\gamma}$	$\hat{\theta}-\theta$	PAIR	$\hat{\gamma}$	$\hat{\theta}-\theta$	PAIR	$\hat{\gamma}$	$\hat{\theta}-\theta$
SHORT-	Uni-	-2.397	0.000	Uni-	-1.790	-0.051	Tri-	-1.199	-0.004
-SHORT	-Uni	0.032	0.005	-Tri	0.052	0.262	-Tri	0.066	0.061
		-2.283	0.017		-1.619	0.668		-1.068	0.158
		-2.462	-0.014		-1.933	-0.888		-1.421	-0.237
SHORT-	Uni-	1.795	-0.595	Uni-	4.809	-0.186	Tri-	5.353	-0.036
-LONG	-Lap	0.677	1.760	-Exp	1.202	0.601	-Exp	1.126	0.328
		6.593	3.487		8.869	1.083		10.807	0.824
		0.886	-4.801		2.605	-1.813		3.391	-0.719
LONG-	Lap-	5.900	-0.006	Exp-	8.915	-0.117	Exp-	6.087	-0.003
-LONG	-Lap	0.595	0.195	-Lap	1.599	0.339	-Ray	1.181	0.402
		7.089	0.548		16.622	0.619		11.570	0.899
		4.684	-1.168		6.031	-1.212		4.205	-1.405

Table 3.1: Results of the simulations with the EML estimator. Signals are composed of 5000 i.i.d. samples. Each four-element column displays, from top to bottom, the mean, the standard deviation, the maximum and the minimum value obtained for the corresponding parameter and signal combination over 100 Monte Carlo runs. Angle values $(\hat{\theta}-\theta)$ are expressed in degrees.

out to be just 0.195^o, clearly below that limit. This outcome confirms the initial suspicion that the conditions of the Gram-Charlier expansion are too restrictive and, indeed, not necessary to be met if estimator (3.19) is to be applied.

For the sake of comparison with the AML estimator (3.19), and in order to illustrate the comments made in Sections 3.3.3 and 3.3.4, Table 3.2 shows the angle error results obtained by both estimators in a further simulation. This time the two independent sources are composed of 1000 samples drawn from a uniform distribution. The simulation is run under the conditions cited earlier, generating independent source-realizations over 100 MC iterations but feeding the same data into both methods at each iteration. For the AML method, the two cases described at the beginning of Section 3.3.3 are considered: when the sign of the arctangent arguments are not taken into account ("AML-a") and when they are ("AML-b"). The first row corresponds to a true rotation angle of $\theta = 20^o$. As predicted in Section 3.3.3, the outcome is identical for EML and AML-a, because $|\theta| < \frac{\pi}{8}$, but a -45^o bias appears in the results of AML-b, since the sources have negative kurtosis and $\theta > 0$. The second row of Table 3.2 shows the results for a new independent set of source realizations and a new angle, $\theta = -30^o < -22.5^o$. Note that now an exact 45^o bias comes up in the results of both AML estimators, just as anticipated in Section 3.3.3: in AML-a because $|\theta| > \frac{\pi}{8}$, and in AML-b because the sks is negative and $\theta < 0$. The final column summarizes the performance under the same source realizations of the closed-form formula for the estimation of θ found by Comon in [41]. EML results are consistently more reliable than those offered by this latter analytical expression. Comon's estimator will be studied in detail in Section 3.4, where this preliminary result will be theoretically endorsed.

Performance variation with sks

We are now concerned with the EML performance as a function of the sks. In order to allow a gradual variation of this parameter, we resort to the ideas of [135], which are also summarized in Appendix B. In short, a signal of any normalized kurtosis value can be modelled by means of a pseudorandom binary sequence (PRBS) and a suitable choice of the probability of its two events. Exploiting this notion, the kurtosis of one of the sources is varied between -2 and 14, whereas the other is kept fixed at

METHOD / ANGLE	EML	AML-a	AML-b	CF
$\theta = 20^{\circ}$	0.006	0.006	−44.994	1.382
	0.033	0.033	0.033	1.145
	0.123	0.123	−44.877	5.117
	−0.056	−0.056	−45.056	−1.217
$\theta = -30^{\circ}$	0.002	45.002	45.002	−0.698
	0.034	0.034	0.034	0.614
	0.088	45.088	45.088	0.994
	−0.109	44.891	44.891	−2.527

Table 3.2: Comparison of several estimators: EML, AML without regard to argument signs (AML-a), AML considering argument signs (AML-b) and Comon's formula [41] (CF). Sources are composed of 1000 independent uniformly distributed samples. Each four-element column shows the mean, the standard deviation, the maximum and the minimum value obtained for the error $(\hat{\theta} - \theta)$ over 100 Monte Carlo runs. All values are expressed in degrees. Experiments for different angle values are independent.

-2 (corresponding to a binary sequence of equiprobable symbols). Hence, sks values are obtained in between -4, corresponding to two equiprobable-symbol binary signals, and 12, which would simulate two exponentially distributed signals (as far as the sks is concerned). At each kurtosis value, the performance index is calculated from separations performed on orthogonal mixtures of fixed $\theta = 10^{\circ}$ over 100 MC runs.[6] The index is the ISR^2 in dB. The solid line in Figure 3.7 shows the results for $T = 5 \cdot 10^3$. The dashed line corresponds to the predicted variance (3.69), which again is a very good approximation of the empirical outcome over all range of sks. The approximation quality becomes even more precise as T increases, as the dotted line (with '$\times$' markers) illustrates for $T = 10^4$. Note the performance degradation around $\gamma = 0$, just as expected (Section 3.3.5). However, due to the finite sample size (and, hence, existing cumulant estimation errors) the performance deterioration is not so severe as theoretically predicted. Also, remark the deterministic performance (perfect estimation) at $\gamma = -4$, i.e., when equiprobable binary sources are processed.

[6] The exact value used for θ is irrelevant, according to the proven orthogonal invariance of the EML estimator.

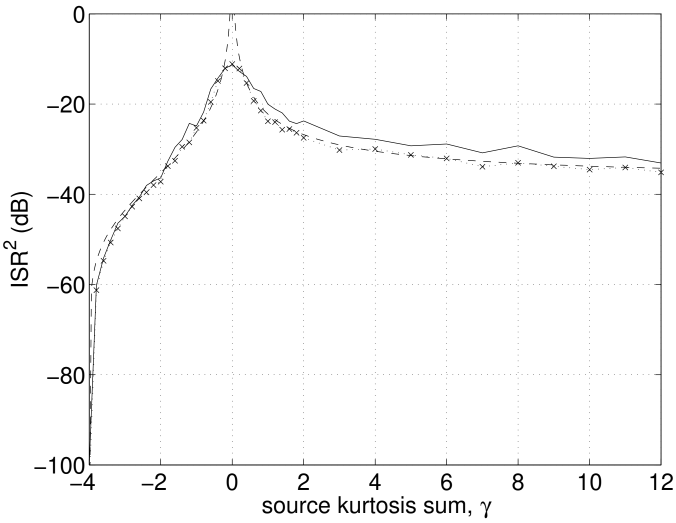


Figure 3.7: Performance of the EML estimator as a function of the source kurtosis sum, for PRBS sources. 100 MC runs per sks value. Solid line: empirical values of ISR for $T = 5 \cdot 10^3$ samples/signal. Dotted line (with '×' markers): empirical values of ISR for $T = 10^4$ samples/signal. Dashed line: variance of $\hat{\theta}$ as predicted by eqn. (3.69) ($10 \log_{10} \sigma^2_{\mathrm{EML}}$).

Validation of analytic asymptotic pdf and variance

Now, we test the validity and accuracy of the analytical asymptotic expression obtained for the pdf of the EML error $\Delta\hat{\theta}$. We compare the empirical and the expected analytic results for the pdf $p_{\Delta\hat{\theta}}$. The source distributions are chosen to be the exponential and uniform distributions. Two sample sizes are tested: $T = 10^3$ and $T = 5 \cdot 10^3$. From each mixture realization, another realization of the error $\Delta\hat{\theta}$ is calculated as $\Delta\hat{\theta} = \hat{\theta}_{\mathrm{EML}} - \theta$, where $\theta = 20^{\mathrm{o}}$ and the EML estimates are obtained from (3.32) and the cumulant expressions (3.34) and (3.35). From 1000 realizations of $\Delta\hat{\theta}$, its pdf is estimated by using a standard method, the naive pdf estimator [115], with window width h. This pdf estimate corresponds to the solid curves in Figure 3.8. Parameters γ, σ_1 and σ_2 are calculated from the source population statistics with the aid of Tables A.1 and A.2. Substituting these coefficients in (3.66) and bearing in mind relationship (3.50), one obtains the analytic pdf estimate represented by the dotted lines in Fig. 3.8. Similarly, the dashed lines correspond to the Gaussian approximation (3.67). Note how all curves become more similar as T increases, thus validating both the asymptotic pdf

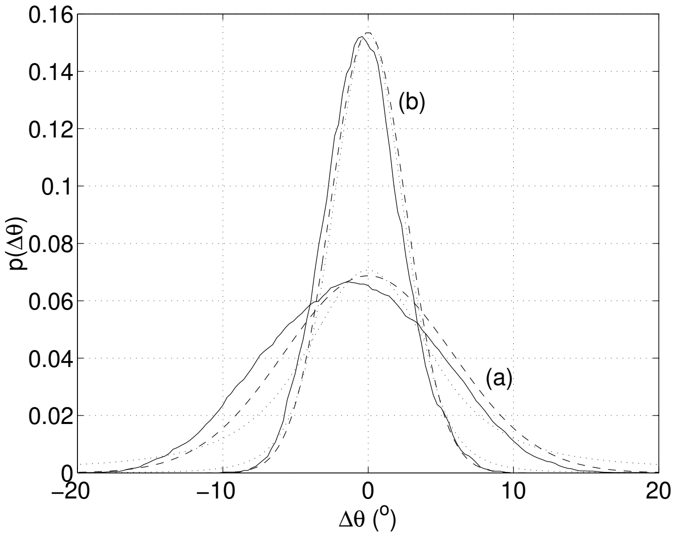


Figure 3.8: Pdf of the EML error for exponential and uniform sources, 1000 signal realizations. Solid lines: naive density estimate. Dotted lines: analytic asymptotic pdf (3.66). Dashed lines: Gaussian approximation (3.67). (a) $T = 10^3$ samples/signal, naive pdf estimate window width $h = 10^o$; (b) $T = 5 \cdot 10^3$ samples/signal, naive pdf estimate window width $h = 3.53^o$.

(3.66) and its Gaussian approximation (3.67). The approximation between estimated and analytical pdf's is even more precise if the parameters of the analytical pdf's are evaluated from the data, instead of using the population values.

On the other hand, Figure 3.9 plots the MSE of the EML estimates versus the number of samples per signal, for the above source combination. At each sample-size value T, ν MC iterations are run, with $\nu T = 5 \cdot 10^5$. The predicted asymptotic variance (3.69) becomes a better approximation as the sample size increases.

A few other examples illustrating and verifying the asymptotic results are detailed in [139].

Influence of pre-whitening

All the experiments reported up to now assess the estimator performance when only the higher-order processing part of the separation is considered. In a practical separation system, however, the EML estimator (3.32) would have to be applied after the sensor output has been whitened. Intuitively, the estimator performance when second-order

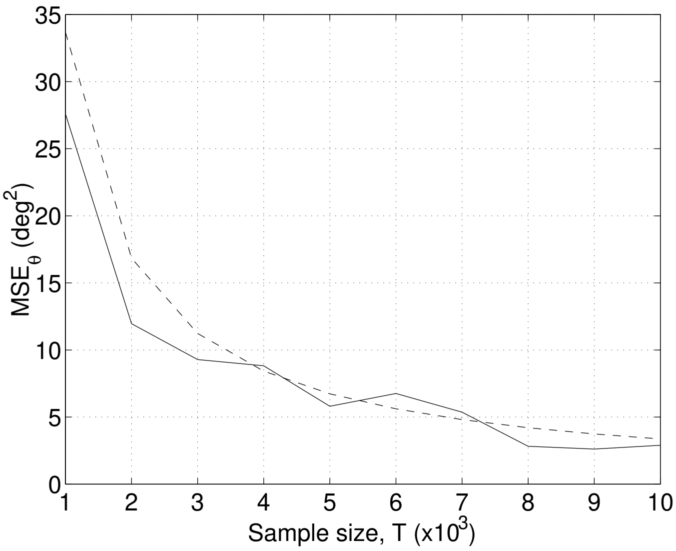


Figure 3.9: Mean squared error of the EML estimates as a function of the sample size, for exponential–uniform sources. Solid line: empirical values, ν Monte Carlo iterations, with $\nu T = 5 \cdot 10^5$. Dashed line: variance from analytic expression (3.69).

processing is included would differ from the results when it is not, since the uncertainties generated by the pre-whitening would add to those introduced by the EML estimator on its own.

In order to empirically check this claim, one can compare the variance of the angle estimates obtained with and without pre-whitening for fixed mixing matrices. Figure 3.10 shows how these variances behave as the sample size increases, for two uniformly distributed sources. The solid line represents the variance of the EML angle estimates for a 30^{o}-rotation on the sources. The dash-dotted line (with '$\times$' markers) includes also the pre-whitening applied on the same sensor-output realizations, followed by EML estimation. The dash-dotted line (with no markers) corresponds to a particular mixing matrix with non-orthogonal transfer vectors, with second plus fourth order separation stages. The pre-whitening process is accomplished by means of the SVD of the sensor-output data matrix, as explained in Section 2.2.1. The variances at each sample-size value are obtained by sample averaging over ν MC runs, with $\nu T = 5 \cdot 10^5$. The same source realizations are used in all three cases. As expected for i.i.d. signals, the variances decrease as T^{-1}. This is endorsed by the dotted lines of Fig. 3.10, which represent least-squares approximations to the curves of the form 'constant/T'. The results for the first and second curves are identical, whereas for the third are slightly higher. Second-order processing introduces errors in the estimation of the whitening matrix which translate into inaccuracies in the resulting orthogonal rotation operating on the sources. Consequently, the variance of the angle estimates increases when pre-whitening is incorporated into the separation system. These results also suggest that the variance increase seems to depend on the mixing matrix conditioning. Nonetheless, further simulations show that this dependence is not too acute, but the variance trajectory remains essentially the same, and very close to that without pre-whitening, over a wide range of mixing matrix conditioning. In particular, when the mixing matrix is orthogonal, the effects of the second-order processing stage on the performance of the EML estimator are totally negligible.

The dashed line of Fig. 3.10 shows the variance calculated from asymptotic expression (3.69). Note how accurate this prediction is with this type of sources, even for low T.

3.4 Comon's Direct Solution

By nulling a 4th-order cross-cumulant of the signals at the separator output, a closed-form expression for the estimation of angle θ was found in [41, 42]. This formula reads:

$$\operatorname{tg} \hat{\theta}_{\mathrm{CF}} = -\rho/2 + \operatorname{sign}(\rho)\sqrt{\rho^2/4+1}, \tag{3.73}$$

where ρ was defined in eqn. (3.38):

$$\rho \triangleq \frac{\kappa_{31}^z - \kappa_{13}^z}{\kappa_{22}^z}. \tag{3.74}$$

The acronym *CF* stands for *Comon's (direct) formula*. In this section we characterize this estimator in more detail and relate it to the estimators studied in the first part of this chapter.

In the first place, we will see next that there is an alternative path, with no explicit cross-cumulant zeroing at the separator output, that leads to an equivalent expression [53, 80] (Section 3.4.1). This alternative expression will prove very useful when studying the CF asymptotic performance in Section 3.4.2. That section also discloses

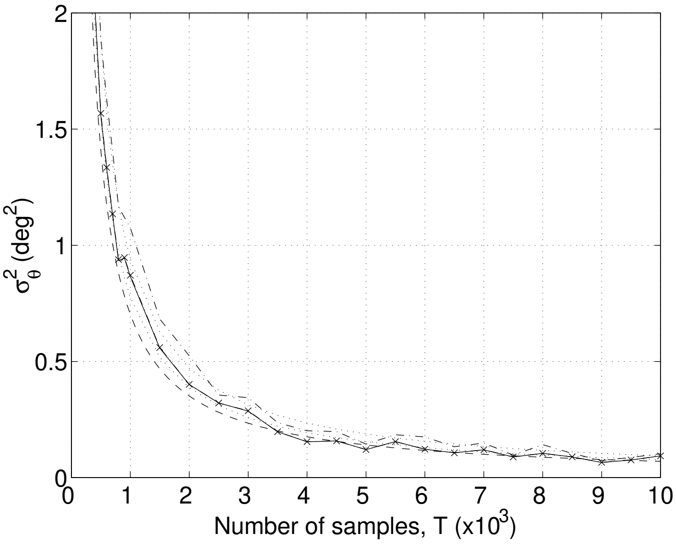


Figure 3.10: Influence of pre-whitening on the variance of the EML estimator, for two i.i.d. sources with uniform distribution. Solid line: EML estimation only, orthogonal mixing matrix. Dash-dotted line with '×' markers: pre-whitening plus EML estimation, orthogonal mixing matrix. Dash-dotted line, no markers: pre-whitening plus EML estimation, non-orthogonal mixing matrix. Dotted lines: $\mathrm{O}(T^{-1})$ LS approximations. Dashed line: asymptotic variance (3.69).

the previously unnoticed bonds between the CF and other two direct expressions for BSS in real mixtures, namely, the AML and the EML.

3.4.1 Alternative development

From eqns. (3.11), parameter ρ in (3.74) can be restated as:

$$\rho = \frac{\kappa_{31}^z - \kappa_{13}^z}{\kappa_{22}^z} = \frac{(\cos^3\theta\,\sin\theta - \cos\theta\,\sin^3\theta)\,(\kappa_{40}^x + \kappa_{04}^x)}{\cos^2\theta\sin^2\theta\,(\kappa_{40}^x + \kappa_{04}^x)}, \tag{3.75}$$

which, eliminating the sks and expanding the trigonometric functions, results in $\rho = 2\cos(2\theta)/\sin(2\theta)$. Therefore, θ may be estimated through [53, 80]:

$$\mathrm{tg}(2\hat{\theta}_{\mathrm{ACF}}) = \frac{2}{\rho}. \tag{3.76}$$

Next it is shown that (3.73) and (3.76) are indeed equivalent. This is why the latter expression may be referred to as *alternative Comon's formula (ACF)*.

In order to prove the equivalence between estimators (3.73) and (3.76), we resort to the formula of the tangent of the double angle:

$$\mathrm{tg}(2\theta) = \frac{2\,\mathrm{tg}\,\theta}{1 - \mathrm{tg}^2\theta}. \tag{3.77}$$

Then, $\mathrm{tg}\,\theta$ can be written as a function of $\mathrm{tg}(2\theta)$ by solving the corresponding second-order equation, yielding:

$$\mathrm{tg}\,\theta = \frac{-1}{\mathrm{tg}(2\theta)} \pm \sqrt{\frac{1}{\mathrm{tg}^2(2\theta)} + 1}. \tag{3.78}$$

Substituting above the value found in (3.76) for $\mathrm{tg}(2\hat{\theta}_{\mathrm{ACF}})$, we obtain:

$$\mathrm{tg}\,\hat{\theta}_{\mathrm{ACF}} = -\rho/2 \pm \sqrt{\rho^2/4 + 1}, \tag{3.79}$$

Now, for (3.79) to provide angle values in the interval $\Omega_{\pi/2} = [-\pi/4, \pi/4]$, the sign of the square root must coincide with $\mathrm{sign}(\rho)$, which transforms (3.79) into (3.73). Equivalently, one can also arrive at (3.76) by substituting eqn. (3.73) into (3.77). Hence, *$\hat{\theta}_{\mathrm{CF}}$ and $\hat{\theta}_{\mathrm{ACF}}$ are simply the same estimates.*

3.4.2 Link with AML and EML estimators. Discussion

Application of (3.77) on $2\hat{\theta}$, together with relationship (3.76) produce:

$$\operatorname{tg}(4\hat{\theta}) = \frac{2\operatorname{tg}(2\hat{\theta})}{1-\operatorname{tg}^2(2\hat{\theta})} = \frac{4\rho}{\rho^2-4}, \tag{3.80}$$

Using cumulant relationship (3.38) on the above equation yields:

$$\operatorname{tg}(4\hat{\theta}) = \frac{4\rho}{\sigma-6} = \frac{4(\kappa_{31}^z - \kappa_{13}^z)}{\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z}. \tag{3.81}$$

The last equality coincides with the approximate AML estimator (3.19). Remark that, as opposed to the development in [69], this expression has been reached without making any assumption on the source statistics. The only condition to apply (3.19) is that the sks must be different from zero. Effectively, we may write (3.19) as the argument (angle with respect the real axis) of ξ in (3.21)/(3.34), whereby it becomes patent that no meaningful estimation of θ can be achieved if $\kappa_{40}^x + \kappa_{04}^x = 0$. Recall also that (3.19) manifests a $\pm 45^o$ bias when the sks is negative (if the signs of the arctangent arguments are considered, Section 3.3.3). Indeed, that estimator was developed under the assumption of sources with positive kurtosis, but we want to stress again that only the sks must be positive, which is certainly a weaker assumption. The positive-sks shortcoming is surmounted if the estimation is explicitly carried out as a function of parameter ρ, as in equation (3.80). This is equivalent to dividing the real and imaginary parts of centroid (3.34) by κ_{22}^z, which, according to (3.11c), has always the same sign as the sks. That is, combining (3.11c) and (3.34), we have:

$$\frac{(\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z)}{\kappa_{22}^z} + j\frac{4(\kappa_{31}^z - \kappa_{13}^z)}{\kappa_{22}^z} = (\rho^2-4) + j4\rho = (\cos\theta\,\sin\theta)^{-2}\,\mathrm{e}^{j4\theta}. \tag{3.82}$$

Accordingly, estimator (3.73)/(3.76) is valid for any sign of sks. However, this advantage occurs at the expense of a dependency of the estimator performance on the unknown parameter θ, as we prove in the next paragraph. But first, it must be remarked that for finite sample size, relation (3.38) does not hold any more, and hence one cannot arrive at (3.80) from (3.73) or (3.76). Only (3.73) and (3.76) are identical even for finite sample size. For infinite sample length (and hence exact cumulant values) all methods provide, obviously, the same solution: the true value of the unknown parameter.

To understand why the CF dependency occurs, we proceed to study such estimator by adopting the strategy employed in Section 3.3.5 to analyze the asymptotic performance of the EML. Having seen the equivalence between (3.73) and (3.76), we decide to focus on the latter, since $\hat{\theta}_{\text{ACF}}$ can directly be written as a function of the argument of the complex quantity:

$$\xi_{\text{ACF}} = (\kappa_{31}^{z} - \kappa_{13}^{z}) + j2\kappa_{22}^{z} = \frac{1}{2}\sin 2\theta\,\gamma\,\mathrm{e}^{j2\theta}. \tag{3.83}$$

In practice, given a pair of whitened signals composed of T samples, $\{\mathbf{z}(k),\ k = 1, \ldots, T\}$, the exact cumulant values would have to be replaced by some cumulant estimates. The performance of a particular estimator is hence determined by its sampling distribution [34, Section IV], [120]. Assuming unit-variance uncorrelated whitened observations, and using the sample estimates of the moments from which the cumulants may be obtained [120] [see also eqns. (3.71)], this would result in the estimate:

$$\hat{\xi}_{\text{ACF}} = \frac{1}{T}\left\{\sum_{k=1}^{T}\big(z_1(k)^3 z_1(k) - z_1(k)z_2(k)^3\big) + j2\sum_{k=1}^{T}\big(z_1^2(k)z_2^2(k) - 1\big)\right\}. \tag{3.84}$$

Expanding this expression as a function of the sources and the unknown parameter [via (3.1)–(3.3)], and after some algebraic manipulations, one arrives at:

$$\hat{\xi}_{\text{ACF}} = \mathrm{e}^{j2\theta}\Delta\hat{\xi}_{\text{ACF}}(\mathbf{x},\,\theta), \tag{3.85}$$

with

$$\Delta\hat{\xi}_{\text{ACF}}(\mathbf{x},\,\theta) \triangleq \frac{1}{4jT}\sum_{k=1}^{T}\Big\{\big(x_1(k) + jx_2(k)\big)^4\mathrm{e}^{j2\theta} - \big[\big(x_1^2(k) + x_2^2(k)\big)^2 - 8\big]\mathrm{e}^{-j2\theta}\Big\}. \tag{3.86}$$

The error committed by $\hat{\theta}_{\text{ACF}}$ in the estimation of θ through $\hat{\xi}_{\text{ACF}}$ is then:

$$\Delta\hat{\theta}_{\text{ACF}} \triangleq \hat{\theta}_{\text{ACF}} - \theta = \frac{1}{2}\arg\big(\Delta\hat{\xi}_{\text{ACF}}(\mathbf{x},\,\theta)\big). \tag{3.87}$$

From the last two equations it now becomes apparent that the performance of the CF estimator is dependent on the parameter to be estimated θ, an undesirable deficiency which is not shared by the AML and the EML estimators. More details follow next.

3.4.3 Asymptotic performance

Given the complex variable

$$\hat{\xi}_x = \omega_1 + j\omega_2 = \frac{1}{T}\sum_{k=1}^{T}\big(\tilde{\omega}_1(k) + j\tilde{\omega}_2(k)\big), \tag{3.88}$$

such that $\tilde{\omega}_1(k)$ and $\tilde{\omega}_2(k)$ are functions of the i.i.d. random vector $\mathbf{x}(k)$, and

$$\begin{aligned} m &= \mathrm{E}[\omega_1] = \mathrm{E}[\tilde{\omega}_1], & \sigma_1^2 &= \mathrm{Var}[\omega_1] = \frac{1}{T}\mathrm{Var}[\tilde{\omega}_1] \\ \mathrm{E}[\omega_2] &= \mathrm{E}[\tilde{\omega}_2] = 0, & \sigma_2^2 &= \mathrm{Var}[\omega_2] = \frac{1}{T}\mathrm{Var}[\tilde{\omega}_2] \end{aligned} \tag{3.89}$$

then the asymptotic pdf (i.e., as $T \to \infty$) of $\delta = \angle\hat{\xi}_x$ is obtained by substituting m for γ in (3.66). In Section 3.3.5, parameters (3.89) were obtained for the EML estimator, for which $\hat{\xi}_x$ is given by (3.46). In the case of the CF estimator, the associated $\hat{\xi}_x$ is given by (3.86) and, from (3.87), $\Delta\hat{\theta}_{\mathrm{ACF}} = \delta/2$. Some tedious but straightforward algebraic simplifications lead to:

$$\mathrm{E}[\tilde{\omega}_1] = m = \frac{1}{2}\left(\kappa_{40}^x + \kappa_{04}^x\right)\sin 2\theta = \frac{1}{2}\left(\mu_{40}^x + \mu_{04}^x - 6\right)\sin 2\theta \tag{3.90}$$

$$\mathrm{E}[\tilde{\omega}_2] = 0 \tag{3.91}$$

$$\begin{aligned} \mathrm{E}[\tilde{\omega}_1^2] = {} & (\mu_{60}^x + \mu_{06}^x - 2\mu_{40}^x\mu_{04}^x)\cos^2 2\theta + \\ & + \frac{1}{4}\left[\mu_{80}^x + \mu_{08}^x + 6\mu_{40}^x\mu_{04}^x - 4(\mu_{60}^x + \mu_{06}^x) - 8(\mu_{40}^x + \mu_{04}^x) + 32\right]\sin^2 2\theta - \\ & - 3(\mu_{50}^x\mu_{03}^x - \mu_{30}^x\mu_{05}^x)\sin 2\theta\cos 2\theta \end{aligned} \tag{3.92}$$

$$\begin{aligned} \mathrm{E}[\tilde{\omega}_2^2] = \tilde{\sigma}_2^2 = {} & 4(\mu_{40}^x\mu_{04}^x - 1)\cos^2 2\theta + (\mu_{60}^x + \mu_{06}^x - 2\mu_{40}^x\mu_{04}^x)\sin^2 2\theta + \\ & + 4(\mu_{50}^x\mu_{03}^x - \mu_{30}^x\mu_{05}^x)\cos 2\theta\sin 2\theta, \end{aligned} \tag{3.93}$$

from which parameters (3.89) can be derived for the CF estimator. The dependence of its asymptotic performance on θ is clearly evidenced in the above expressions. The asymptotic properties of the EML were studied in Section 3.3.5. Similarly, the CF is also *asymptotically unbiased* and *strongly consistent* provided $m \neq 0$, that is, as long as both the sks *and* θ are not null.

Gaussian approximation (3.67) also holds for the CF estimator, for its corresponding parameters. Accordingly, its asymptotic variance is given by:

$$\mathrm{Var}[\hat{\theta}_{\mathrm{CF}}] \xrightarrow[T\to\infty]{} \sigma_{\mathrm{CF}}^2 = \frac{\tilde{\sigma}_2^2}{T(2m)^2}, \tag{3.94}$$

where m and $\tilde{\sigma}_2^2$ appear in (3.90) and (3.93), respectively. Developing the above expression and taking into account (3.69), one reaches:

$$\sigma_{\mathrm{CF}}^2 = \sigma_{\mathrm{EML}}^2 + \Delta\sigma^2(\theta), \tag{3.95}$$

in which,

$$\Delta\sigma^2(\theta) \triangleq \frac{4}{T\gamma^2 \mathrm{tg}\, 2\theta} \left\{ \frac{(\mu_{40}^x \mu_{04}^x - 1)}{\mathrm{tg}\, 2\theta} + (\mu_{50}^x \mu_{03}^x - \mu_{30}^x \mu_{05}^x) \right\} \tag{3.96}$$

is the *excess variance* between the CF and the EML. Expressions (3.95)–(3.96) are remarkable in two aspects. First, they clearly show that the asymptotic performance of the CF estimator depends on the true parameter θ. Second, they formulate the specific linkage between the EML and the CF asymptotic variances. Note that when at least one of the sources is symmetric or both sources have identical distribution (actually, it is only required that they have equal 3rd and 5th order moments), the excess variance is positive or null for all values of θ, indicating that $\sigma_{\mathrm{CF}}^2 \geqslant \sigma_{\mathrm{EML}}^2$, $\forall\theta$. When both sources are asymmetric and do not share the same distribution, there could exist, in principle, values of θ for which $\sigma_{\mathrm{CF}}^2 < \sigma_{\mathrm{EML}}^2$. Also, in such a case an asymmetric performance with θ could be manifested by the CF, as will be illustrated in Section 3.4.5. For low θ the CF will generally show considerably higher variance than the EML.

3.4.4 Computational complexity

Both equivalent forms (3.73) and (3.76) require the computation of the whitened 4th-order cross-cumulants κ_{31}^z, κ_{22}^z and κ_{13}^z, but the ACF presents the advantage over the CF that the square rooting is avoided. Proceeding as in Section 3.3.6, the cumulant computation takes, for T sufficiently high, around $6T$ flops per signal pair. This can be further reduced to $5T$ flops if the common term $z_1 z_2$ is factorized from the sample estimate of $(\kappa_{31}^z - \kappa_{13}^z)$ [see the real part of eqn. (3.84)].

3.4.5 Simulations

The following experiments evaluate the performance of the CF estimator with respect to angle θ and the sks γ, and check the accuracy of the asymptotic results.

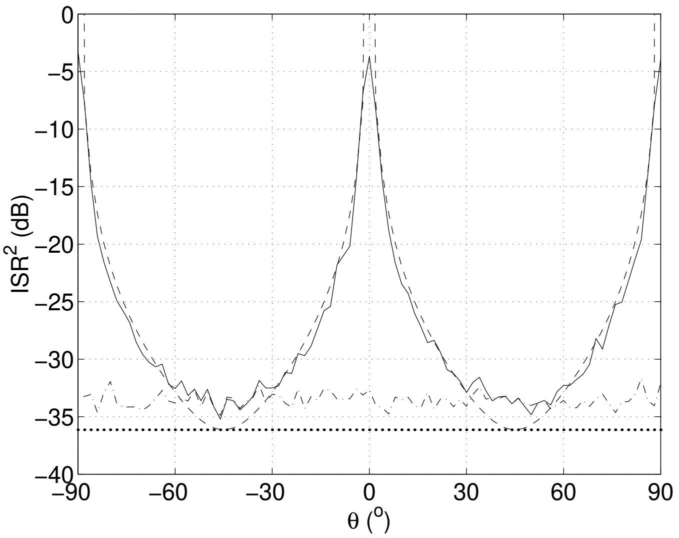
CF's behaviour vs. θ

Figure 3.11(a) shows the variation of the CF performance as a function of the true angle θ (solid line), for 5000-sample uniform and asymmetric triangular (the short-skewed distribution presented in Section 3.3.7) source distributions. Equation (3.73) was used. The asymptotic variance (3.94) is a very precise approximation to the empirical results (dashed line), corroborating the quality of the asymptotic analysis of Section 3.4.3. For the sake of comparison, the EML results for the same experiment are also represented (dash-dotted and dotted lines). The excess variance (3.96) was also validated by checking that it was very close to the difference between the CF and the EML ISR results over the whole angle range.

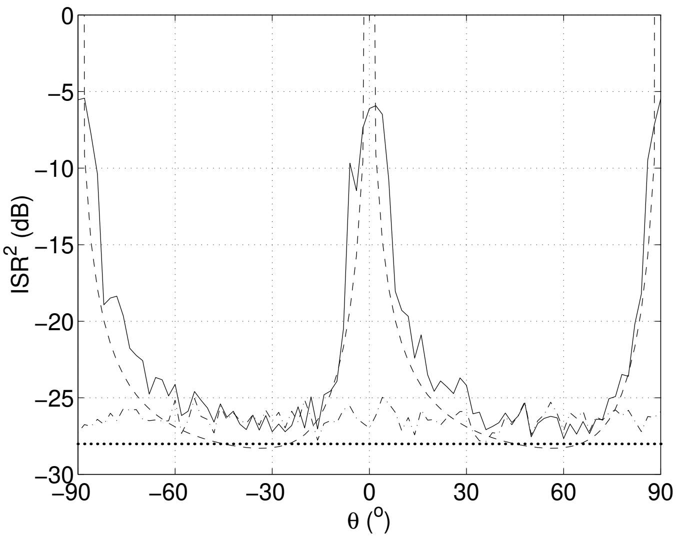
To illustrate the comments made at the end of Section 3.4.3, the same experiment is conducted, but this time the uniform source is changed to an asymmetric pdf: an exponential distribution, so that the sources have different asymmetric pdf's. The results are plotted in Figure 3.11(b). Note that now the CF performance is not an even function of θ any more. Also, the asymptotic performance of the CF improves slightly that of the EML near $\theta \equiv \pi/4$, although this difference is not significant in the empirical outcomes.

Variation with sks

The performance variation of the CF with the sks is checked next. Again, PRBS sources are used, as in the analogous experiments of Section 3.3.7. The solid line in Figure 3.12 shows the MSE of (3.73) obtained for $T = 5 \cdot 10^3$ samples/signal over 100 MC runs at each sks value, and $\theta = 10^{\mathrm{o}}$. The dashed line corresponds to the predicted asymptotic variance, evaluated from (3.94) and the normalized moments in Table A.1. The similarity between the empirical and the theoretical values becomes even closer for higher sample size, as shown by the dotted line, obtained for $T = 10^4$ samples/signal. Similarly as occurred for the EML, a quality degradation is observed near $\gamma = 0$, as anticipated from the analysis in the previous sections.



(a)



(b)

Figure 3.11: Performance of the CF estimator as a function of the true angle θ, $T = 5 \cdot 10^3$ samples/signal, for (a) uniform and asymmetric triangular, and (b) exponential and asymmetric triangular source distributions. Solid and dash-dotted lines: empirical ISR^2 for the CF and EML, respectively, averaged over 100 MC runs per angle value. Dashed and dotted lines: variance of $\hat{\theta}_{\mathrm{CF}}$ and $\hat{\theta}_{\mathrm{EML}}$ ($10 \log_{10} \sigma^2$) predicted from (3.94) and (3.69), respectively.

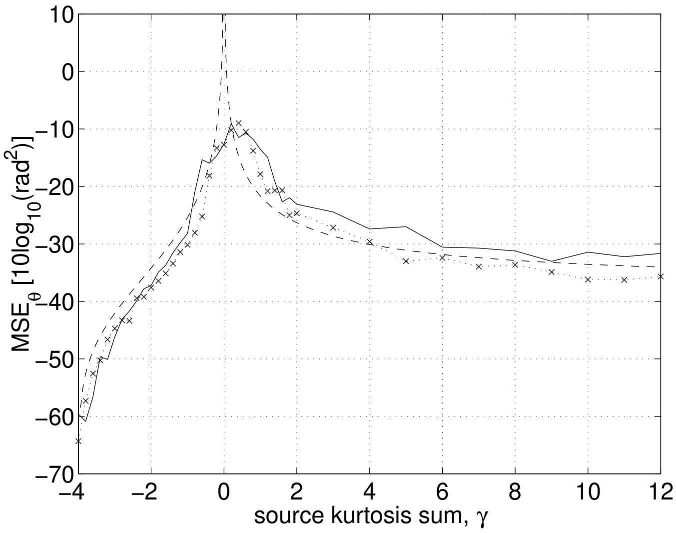


Figure 3.12: Performance of the CF estimator as a function of the sks, for PRBS sources, $\theta = 10^{\rm o}$, 100 MC runs. Solid line: empirical MSE of $\hat{\theta}_{\rm CF}$ for $T = 5 \cdot 10^3$ samples/signal. Dotted line (with '×' markers): empirical MSE of $\hat{\theta}_{\rm CF}$ for $T = 10^4$. Dashed line: asymptotic variance (3.94).

3.5 Summary and Conclusions

In the first part of this chapter, the AML estimator of [69] has been considered. Due to the validity domain of the Gram-Charlier expansion and certain assumptions made during its development, the derivation of the AML is restricted to symmetric sources with the same distribution and normalized kurtosis lying in the range $[0, 4]$. Along the same lines, a novel expression for the estimation of the rotation angle has been found, also based on the 4th-order statistics of the data. Interestingly, to arrive at the new expression neither pdf expansions nor assumptions on the source pdf's have been necessary. Consequently, under minor modifications the AML is made valid for any source pdf combination, with arbitrary kurtosis value and symmetry. The only actual condition, much weaker than all the previous restrictions, is that the sks must be different from zero. The resulting new expression can be regarded as an extension of the approximate ML solution of [69], and hence the name *extended ML (EML)* estimator.

The new estimator has been related to other direct solutions, and its *modus operandi* has been explained in terms of the scatter-plots of the signals involved. This geometrical rationale behind the method becomes specially apparent when the sources have the same distribution with symmetric pdf. From this perspective, it can be argued that the estimator exploits the specific symmetries existing in the jpdf of independent variables, the sources.

A analytical expression for the asymptotic pdf of the estimator when i.i.d. observations are processed has been obtained. From this formula, it has also been seen how the estimator behaves as some parameters (γ, σ_1 and σ_2) vary. These parameters depend on the higher-order statistics (up to the eighth order) of the source signals. Specifically, the larger are $\frac{|\gamma|}{\sigma_1}$ and $\frac{\sigma_1}{\sigma_2}$, the more precise becomes the estimation. In the i.i.d. case the estimator exhibits strong consistency (except when the sks is null) and asymptotic unbiasedness. We have also shown that when there is a symmetric source, the estimator is unbiased for any sample size. If the sample size is large enough, the analytic asymptotic pdf can be approximated by a Gaussian distribution, from which the asymptotic variance of the estimator has easily been revealed.

Several simulations have verified the validity of the theoretical results, offering satisfactory angle estimation in a wide variety of source pdf combinations, with different tails (i.e., kurtosis value and sign) and symmetries. Additional experiments have confirmed that the analytical solutions found for the estimator asymptotic pdf and variance are able to predict with great accuracy the empirical outcomes. It has also been observed that the incorporation of the pre-whitening stage into the separation system has practically negligible effects of the estimator's performance.

In the second part of the chapter, the connections among the CF [41,42], the AML and the EML have been established. The AML could not be used for negative sks, but thanks to the link evidenced, an equivalent expression for this estimator has been found which overcomes this deficiency. The CF has been statistically characterized by using the tools and strategy adopted for the EML, and its performance has been shown to depend on the angular parameter to be estimated. Hence, as opposed to the EML, the CF estimator does not enjoy the orthogonal invariance property. Therefore, the EML estimator emerges as the best option among those three methods: it is valid for any sign of sks and its performance is independent of the unknown parameter. Simulations corroborate this claim.

In conclusion, a novel 4th-order closed-form expression has been put forward and studied, and, in the light of the analysis carried out and the links evidenced, the real-mixture BSS scenario has been endowed with a unifying framework for its closed-form solutions.

Next chapter goes deeper into the new estimator's characterization by approaching it from the contrast-function optimization standpoint.

Chapter 4

Optimization Criteria

4.1 Introduction

As commented in the literature review of Section 2.5, an increasing number of BSS methods are based on the optimization of the so-called *cost* or *contrast functions*. In the general case, the parameters of the mixture are identified by means of a functional whose extremization is proven to be a sufficient condition for the source extraction. For instance, in the well-known method of [48], the contrast function is the sum of squares of the output 4th-order marginal cumulants. The maximization of this function is shown to supply a valid separating matrix.

Attacking the BSS problem from a contrast-function optimization perspective presents a number of beneficial aspects. In the first place, these types of methods are inherently endowed with some advantageous properties. For instance, the optimization of a function of certain distributional property of the output subject to the output spatially-whiteness constraint leads automatically to equivariant algorithms enjoying the desirable uniform performance feature (see Section 2.3) [24,25]. In the second place, well-established theory exists — such as that of the closely related *estimating functions* and *estimating equations* [30, 65], or that of (optimal) *statistic matching* [26, 28] — on which methods of this sort can capitalize in order to develop further results. In the third place, the fact that an algorithm is associated with a contrast function may facilitate the analysis of the existing bonds with other algorithms also based on analogous principles. Also, departing from a cost function an adaptive separation algorithm can easily be devised through the corresponding stochastic-gradient updating scheme [11, 95] (see Chapter 7).

In this chapter, the EML estimator is seen as resulting from the optimization of certain cumulant matching and contrast function criterion (Section 4.2). From this standpoint, a number of additional results on the estimator are obtained. First, we establish an interesting connection between the EML estimator and another contrast-based method encountered in the literature (Section 4.3). Further, Section 4.4 shows that the maximization of the whitened-observations log-likelihood under the conditions of the Gram-Charlier pdf expansion and sources with the same symmetric distribution is equivalent to the EML contrast function, thus confirming the relationship between the EML and the AML estimators exposed in the preceding chapter. Finally, a small-error analysis of the EML contrast function (Section 4.5) corroborates the asymptotic results obtained in Section 3.3.5.

4.2 Cumulant Matching and Contrast Function Optimization Criteria

From this point on, we will adopt a more formal approach when presenting results of a mathematical nature. The benefits are twofold. First, it is actually the clearest possible manner to present information of this sort. Second, it is also the briefest. We start the exposition of this chapter by formalizing some results already encountered in the preceding chapter.

Definition 4.1 (fourth-order centroid). Expression

$$\xi_4(\mathbf{z}) \triangleq \mathrm{E}[(z_1 + jz_2)^4], \tag{4.1}$$

is called the *fourth-order centroid (FOC)* of the whitened signals. Analogous terminology holds for the other elements of the separation system (sources and outputs).

For zero-mean unit-variance components, the FOC is easily expanded as a function of the 4th-order cumulants:

$$\xi_4(\mathbf{z}) = \big(\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z\big) + j4\big(\kappa_{31}^z - \kappa_{13}^z\big). \tag{4.2}$$

And, accordingly:

$$\xi_4(\mathbf{x}) = \kappa_{40}^x + \kappa_{04}^x = \gamma, \tag{4.3}$$

which is the sks, as defined in (3.25). On the other hand, due to the scatter-plot relationship (3.14), the connection between the FOC's can be formalized in the following manner:

Lemma 4.2 (centroid rotation). If $\mathbf{z} = \mathcal{Q}(\theta)\mathbf{x}$, then $\xi_4(\mathbf{z}) = \mathrm{e}^{j4\theta}\xi_4(\mathbf{x})$. Consequently, if $\mathbf{s} = \mathcal{Q}(-\hat{\theta})\mathbf{z} = \mathcal{Q}(\theta - \hat{\theta})\mathbf{x}$, then $\xi_4(\mathbf{s}) = \mathrm{e}^{j4(\theta-\hat{\theta})}\xi_4(\mathbf{x})$ and $\xi_4(\mathbf{s}) = \mathrm{e}^{-j4\hat{\theta}}\xi_4(\mathbf{z})$. In particular: $|\xi_4(\mathbf{s})| = |\xi_4(\mathbf{z})| = |\xi_4(\mathbf{x})|$, i.e., the centroid modulus is not altered by a rotation.

With these simple premises in mind, we are ready to prove the following result:

Theorem 4.3 (EML optimization criteria). *The EML estimator given analytically by equation* (3.32) *corresponds to the local minimizers of*

$$J_{\mathrm{m}4}(\hat{\theta}) = \big|\xi_4(\mathbf{s}) - \xi_4(\mathbf{x})\big|^2 = \big|\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}} - \xi_4(\mathbf{x})\big|^2, \tag{4.4}$$

or, equivalently, the local maximizers of

$$J_{\mathrm{M}4}(\hat{\theta}) = \mathrm{Re}\big(\xi_4(\mathbf{x})^*\xi_4(\mathbf{s})\big) = \mathrm{Re}\big(\xi_4(\mathbf{x})^*\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}}\big), \tag{4.5}$$

that is,

$$\hat{\theta}_{\mathrm{EML}} = \arg\min_{\hat{\theta}} J_{\mathrm{m}4} = \arg\max_{\hat{\theta}} J_{\mathrm{M}4}. \tag{4.6}$$

Proof. $J_{\mathrm{m}4}(\hat{\theta}) = \big(\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}} - \xi_4(\mathbf{x})\big)\big(\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}} - \xi_4(\mathbf{x})\big)^* = |\xi_4(\mathbf{z})|^2 + |\xi_4(\mathbf{x})|^2 - 2\mathrm{Re}\big(\xi_4(\mathbf{x})^*\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}}\big)$, which, by Lemma 4.2, reduces to $2|\xi_4(\mathbf{x})|^2 - 2J_{\mathrm{M}4}(\hat{\theta})$. From (4.3), $|\xi_4(\mathbf{x})| = |\gamma|$, so $|\xi_4(\mathbf{x})|^2$ is constant for a given pair of sources, and then the minimization of $J_{\mathrm{m}4}$ is tantamount to the maximization of $J_{\mathrm{M}4}$. Thus, we only need to show the second part of the theorem.

The first and second derivatives of $J_{\mathrm{M}4}$ are easily obtained by bearing in mind Lemma 4.2:

$$\frac{\partial J_{\mathrm{M}4}}{\partial\hat{\theta}} = 4\mathrm{Im}\big(\xi_4(\mathbf{x})^*\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}}\big), \qquad \frac{\partial^2 J_{\mathrm{M}4}}{\partial\hat{\theta}^2} = -16\mathrm{Re}\big(\xi_4(\mathbf{x})^*\xi_4(\mathbf{z})\,\mathrm{e}^{-j4\hat{\theta}}\big) = -16J_{\mathrm{M}4}. \tag{4.7}$$

The local extrema of J_{M4} are given by the first-order necessary condition:

$$\frac{\partial J_{\mathrm{M4}}}{\partial \hat{\theta}} = 0 \;\Leftrightarrow\; \mathrm{e}^{j\left[\angle(\xi_4(\mathbf{x})^* \xi_4(\mathbf{z})) - 4\hat{\theta}\right]} = \pm 1. \tag{4.8}$$

The local maxima must fulfil, in addition, the second-order necessary condition $\frac{\partial J_{\mathrm{M4}}}{\partial \hat{\theta}} <$ 0, which occurs if and only if $\mathrm{e}^{j[\angle(\xi_4(\mathbf{x})^* \xi_4(\mathbf{z})) - 4\hat{\theta}]} = 1$. Hence $\hat{\theta}_{\mathrm{opt}} = \frac{1}{4}\angle\big(\xi_4(\mathbf{x})^* \xi_4(\mathbf{z})\big) + \frac{n\pi}{2}$, $n \in \mathbb{Z}$, which transforms into $\hat{\theta}_{\mathrm{EML}}$ because $\xi_4(\mathbf{x})$ is real and the term $\frac{n\pi}{2}$ is immaterial to the source separation. ■

Minimization of (4.4) corresponds to a 4th-order cumulant matching criterion, whereby a linear combination of output cumulants, $\xi_4(\mathbf{s})$, is matched (i.e., its mismatch reduced as much as possible) to the same linear combination of source cumulants, $\xi_4(\mathbf{x})$. Maximization of (4.5) yields a cost or contrast function optimization criterion. Through Theorem 4.3, the optimization of these equivalent criteria is solved in closed-form by (3.32). In conclusion, eqns. (4.4)–(4.5) correspond to the optimization criteria associated with the EML estimator. Hence, the question raised in [137] about whether such optimization principle existed for this estimator is now settled. As a by-product, this also proves that the EML estimator is *orthogonal invariant* [24,34] (see also Section 2.3), i.e., its performance is independent of the particular value of Q (property which was already manifested when analyzing its performance in Section 3.3.5).

Though interesting, at first sight this function-optimization result could strike as a pure mathematical curiosity, but, on the contrary, it conveys remarkable consequences. Specifically, by means of these results we are going to extend the applicability of certain contrast function previously proposed in the literature (Section 4.3). But first, let us dwell a little longer on the peculiarities of the cumulant-matching principle (4.4).

4.2.1 Cumulant matching

The statistic-matching approach to the identification of an unknown parameter ϑ (which may well be a vector of parameters in the general case) can be summarized as follows [26,28]. Denote by $\hat{\Gamma}$ a vector of sample statistics of the available data, such that $\mathrm{E}_{\vartheta}[\hat{\Gamma}] = \Gamma(\vartheta)$, where ϑ is the unknown vector paramaterizing the distribution of

$\hat{\Gamma}$. An estimate $\hat{\vartheta}_*$ of ϑ may be obtained as

$$\hat{\vartheta}_* = \arg\min_{\hat{\vartheta}} c(\hat{\vartheta}), \tag{4.9}$$

where $c(\hat{\vartheta})$ is a measure of the discrepancy between $\hat{\Gamma}$ and $\Gamma(\hat{\vartheta})$. The optimal, in the minimum MSE (MMSE) sense, estimate of ϑ corresponds to the cost function:

$$c(\hat{\vartheta}) = \big(\hat{\Gamma} - \Gamma(\hat{\vartheta})\big)^{\mathrm{H}} \mathrm{Cov}^{-1}(\hat{\Gamma})\big(\hat{\Gamma} - \Gamma(\hat{\vartheta})\big), \tag{4.10}$$

matrix $\mathrm{Cov}(\hat{\Gamma})$ being the covariance of the sample statistic $\hat{\Gamma}$. Matrix $\mathrm{Cov}^{-1}(\hat{\Gamma})$ generates the coefficients whereby each statistic mismatch must be weighted in order to obtain optimal MMSE performance.

In the context of BSS, $\hat{\Gamma}$ generally corresponds to a vector of sample cumulants. If the separating matrix is considered as the unknown parameter, and for high enough SNR, thanks to the multiplicative structure of the problem the above criterion can be approximated by:

$$c(\widehat{W}) = \big(\hat{\Gamma}_s - \Gamma_x\big)^{\mathrm{H}} \mathrm{Cov}^{-1}(\hat{\Gamma}_x)\big(\hat{\Gamma}_s - \Gamma_x\big), \tag{4.11}$$

parameter $\widehat{W}$ being an estimate of the separating matrix. Criterion (4.11) has the key feature that the optimal weighting matrix $\mathrm{Cov}^{-1}(\hat{\Gamma}_x)$ does not depend on the mixing matrix itself, but is constant for a given source distribution. Functional $c(\widehat{W})$ depends on $\widehat{W}$ only through the output vector $\mathbf{s} = \widehat{W}\mathbf{y}$. Eqn. (4.11) represents the general form of the so-called *output cumulant matching* criterion for BSS.

The EML estimator as an output cumulant matching criterion

Now, we can identify (4.4) with the above general criterion through the identities: $\Gamma_x = \gamma$ and $\hat{\Gamma}_s = \hat{\xi}_4(\mathbf{s})$ [the sample estimate of $\xi_4(\mathbf{s})$]. Also, in the two-signal case and after pre-whitening, the unknown parameter is not the whole separating matrix, but only the angle θ characterizing the remaining unitary transformation. Parameter γ is the source kurtosis sum, so that the EML estimator may be seen as resulting from minimizing the mismatch between this distributional property of the sources and the corresponding property of the separator output. Minimization of (4.4) is thus a particular output cumulant matching criterion.

As we have just seen, the optimal, in the MMSE sense, estimate of θ would normally require both a prior knowledge of the source cumulants being considered and a consistent estimate of the covariance matrix of the sample statistics of such cumulants [26]. But all this information is unknown in a genuine blind separation problem. Criterion (4.4) surmounts these difficulties. In the first place, the source extraction is accomplished by resorting to a single statistic, the source kurtosis sum, and hence the covariance matrix becomes immaterial in the selection of the optimal weights for the cumulant matching expression. In the second place, the value of the pertinent source statistical parameter ($\Gamma_x = \gamma$) can be directly estimated from the available whitened observations [eqn. (3.27)], and so no prior knowledge on the source distribution is needed. Consequently, eqn. (4.4) constitutes a privileged output cumulant matching criterion.

4.3 Extension of Moreau-Comon's Contrast Function

In this section we extend, in the two-signal case, an existing contrast function. In its original definition, the application of this function required that: a) the kurtosis of all sources have the same sign (same source tail), and b) this sign be known in advance. The EML contrast function, which can be considered as a generalization of the latter, spares these two requirements.

4.3.1 Moreau-Comon's contrast function

In [94] and [95] a class of contrast functions for source separation is defined. It is proved that if all the sources have the same sign of kurtosis, say ε, then:

$$I_4(\mathbf{s}) = \varepsilon \sum_{i=1}^{q} \mathrm{E}[s_i^4], \tag{4.12}$$

is a contrast over the space of spatially white random vectors. When a contrast is to be optimized under the unitary constraint $\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = I$, it is called *orthogonal contrast* [25]. In the two signal case, $q = 2$, orthogonal contrast (4.12) becomes:

$$I_4(\mathbf{s}) = \varepsilon(\mu_{40}^s + \mu_{04}^s), \qquad \varepsilon \triangleq \mathrm{sign}(\kappa_{40}^x) = \mathrm{sign}(\kappa_{04}^x). \tag{4.13}$$

We call (4.13) *Moreau's contrast function (MCF).*

On the other hand, a somewhat cumbersome closed-form solution was given in [52] for the maximization of (4.13). In the sequel, the resulting direct method will be called SOK, for "sum of output kurtosis", since orthogonal contrast (4.12) can equivalently be written as $I_4(\mathbf{s}) = \varepsilon \sum_i \kappa^s_{iiii}$ [52].

4.3.2 EML contrast function. Link

Since $\xi_4(\mathbf{x}) = \gamma$ is real valued, function (4.5) can be expressed as

$$J_{\text{EML}} = J_{\text{M4}} = \gamma \mathbb{R}\text{e}\big(\xi_4(\mathbf{s})\big), \tag{4.14}$$

and also, by virtue of (4.2), as a function of the output cumulants:

$$J_{\text{EML}} = \gamma\,(\kappa^s_{40} - 6\kappa^s_{22} + \kappa^s_{04}). \tag{4.15}$$

Now we prove that the maximization of this function is equivalent to the maximization of another contrast function very similar to (4.13).

Lemma 4.4. The functions J_{EML} in eqn. (4.15) and

$$\tilde{J}_{\text{EML}} = \varepsilon_\gamma\,(\mu^s_{40} + \mu^s_{04}), \qquad \text{with } \varepsilon_\gamma = \text{sign}(\gamma), \tag{4.16}$$

have identical local extrema.

Proof. Due to the relationship between the outputs and the whitened signals, $\mathbf{s} = \mathcal{Q}(-\hat{\theta})\,\mathbf{z}$, it turns out that $\frac{\partial s_1}{\partial \hat{\theta}} = s_2$ and $\frac{\partial s_2}{\partial \hat{\theta}} = -s_1$. By using these equations when differentiating the cumulants, we easily arrive at $J'_{\text{EML}} = 16\gamma\,(\mu^s_{31} - \mu^s_{13})$. Similarly: $\tilde{J}'_{\text{EML}} = 4\varepsilon_\gamma\,(\mu^s_{31} - \mu^s_{13})$. Hence: $J'_{\text{EML}} = 4|\gamma|\tilde{J}'_{\text{EML}}$. Both functions are directly proportional from the first derivative onwards. ■

The consequence of this result is immediate:

Proposition 4.5. *The local maxima of $\tilde{J}_{\text{EML}}$ in eqn.* (4.16) *are given in closed-form by* (3.32).

Therefore the orthogonal contrast (4.16) provides source separation under the same conditions as the EML estimator (3.32). It can be applied to any source distribution, and hence kurtosis value, as long as the sks is not null. Remark, in addition, that contrast function (4.13) becomes (4.16) if the source kurtosis sign ε is substituted by the sign of source kurtosis sum ε_γ. Therefore, in the two dimensional case, by means of (4.16), contrast (4.13) is extended to sources of any kurtosis, including opposite kurtosis sign (providing $\gamma \neq 0$). Moreover, as opposed to (4.13), where one must have a prior knowledge of the source kurtosis sign, the sks sign needs not be known in advance, since it can be evaluated directly from the data [via (3.27)]. This is most desirable in a really blind operation. Finally, the above comments show that, if ε_γ is used instead of ε in (4.13), the awkward SOK solution [52], in the real case, gracefully reduces to $\hat{\theta}_{\mathrm{EML}}$ given by (3.32).

For more than two signals, the pairwise application of contrast $\tilde{J}_{\mathrm{EML}}$ (4.16), i.e., of the EML analytic solution (3.32), provides satisfactory results in all experiments we carried out. This is a somehow surprising outcome, since it is not generally true that the multidimensional extension of $\tilde{J}_{\mathrm{EML}}$ is a valid contrast. The study of the BSS scenario of more than two signals is carried out in Chapter 6.

4.3.3 Empirical validation

The consequences of Proposition 4.5 concerning the relationship between the EML and the SOK are illustrated with a simple computer simulation involving PRBS sources with varying sks. Figure 4.1 shows the results obtained for $T = 5 \cdot 10^3$ samples per signal and 100 MC runs by three angle estimators: EML, SOK and SOK'. The SOK method is, as explained in Section 4.3.1, the direct solution [52] for the maximization of MCF (4.13).[1] It was assumed throughout this simulation that both sources had positive kurtosis, $\varepsilon = 1$. On the other hand, the method labelled as SOK' corresponds to the direct solution SOK, but using the sign of sks ε_γ as estimated by (3.30) [or the sample estimate of the equivalent cumulant relationship (3.35)].

Results of Fig. 4.1 confirm that, in effect, SOK' solutions coincide with the EML solutions. In fact, every single estimate from both methods at each MC iteration were

[1] Original code for the SOK solution was downloaded from the URL:
`http://www.i3s.unice.fr/~comon/matlab.html`.
The original MATLAB function was modified so that instead of calculating the cumulants inside the function, these were entered as input parameters. The purpose of this modification was to make all tested methods work from the same cumulant estimates, thus increasing the significance of subsequent comparisons.

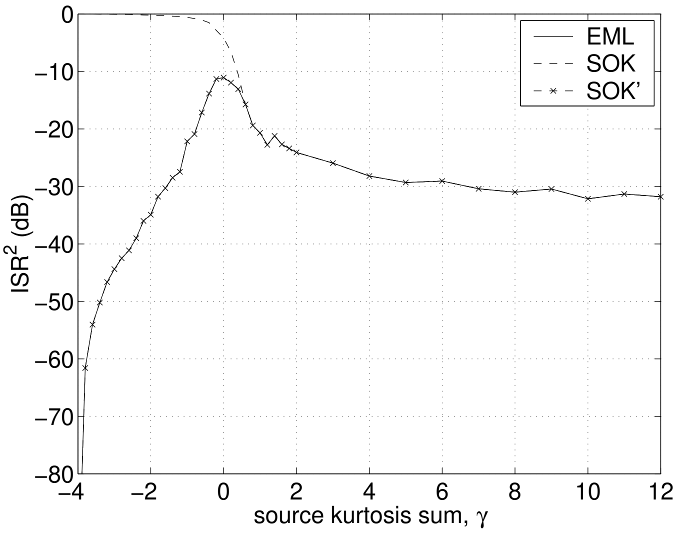


Figure 4.1: Comparative performance of the EML and SOK estimators as a function of the source kurtosis sum, for PRBS sources. $T = 5 \cdot 10^3$ samples/signal, 100 MC runs per sks value.

identical. SOK and EML estimates also coincide when ε and ε_γ are equal (in this experiment, when the sks is positive), even though the source tails are then different, which is against the original MCF conditions. This corroborates that the use of contrast (4.13) can be extended as commented in Section 4.3.2.

4.4 Approximate-Likelihood Maximization

When analyzing the relationship between the AML and the EML estimators in Section 3.3.3, it was claimed that the latter is also the ML estimator of the unknown parameter under the conditions of the AML development (source jpdf Gram-Charlier expansion and symmetric sources with the same positive kurtosis). We now substantiate this statement by proving that the maximization of the EML contrast function (4.14) is effectively equivalent to the maximization of the likelihood function of the whitened observations under the mentioned conditions.

From eqns. (3.15) and (3.18), the approximate log-likelihood function of the whitened observations for sources with identical kurtosis reduces to:

$$\log p_{\mathbf{z}}(\mathbf{z}\,|\,\hat{\theta}) = (\kappa_{40}^x + \kappa_{04}^x) \sum_k \frac{r_k^4}{4!\,8} \cos 4(\phi_k - \hat{\theta}) + k(r), \tag{4.17}$$

in which the term $k(r)$ is rotationally invariant, that is, it does not depend on angle $\hat{\theta}$ but only on the moduli r_k. With the definitions made up to now, this log-likelihood function can be expressed as:

$$\log p_{\mathbf{z}}(\mathbf{z}\,|\,\hat{\theta}) = \frac{T}{4!\,8} \gamma \mathbb{R}\mathrm{e}\big(\hat{\xi}_4(\mathbf{s})\big) + k(r), \tag{4.18}$$

$\hat{\xi}_4(\mathbf{s})$ being the sample estimate of the output FOC $\xi_4(\mathbf{s})$ (Definition 4.1):

$$\hat{\xi}_4(\mathbf{s}) = \frac{1}{T} \sum_k r_k^4 \mathrm{e}^{j4(\phi_k - \hat{\theta})}. \tag{4.19}$$

The derivatives of $\hat{\xi}_4(\mathbf{s})$ with respect to θ are, in turn, the sample estimates of the respective derivatives of $\xi_4(\mathbf{s})$. Consequently, the maximization of the above function is equivalent to the maximization of the sample estimate of the EML contrast function (4.14), leading to the EML estimator (3.32). We conclude that the EML is the ML estimator of θ under the particular constraints of the AML development.

4.5 Small-Error Analysis

In practice, maximization of EML contrast function (4.15) does not provide the exact value of the parameter θ, since the true FOC's are actually approximated by the sample estimates:

$$\hat{\xi}_4(\mathbf{s}) = \frac{1}{T} \sum_{k=1}^{T} \big(s_1(k) + js_2(k)\big)^4 = \frac{1}{T} \sum_{k=1}^{T} \zeta(k) \tag{4.20}$$

$$\hat{\xi}_4(\mathbf{x}) \approx \hat{\gamma} = \frac{1}{T} \sum_{k=1}^{T} \big(z_1^2(k) + z_2^2(k)\big)^2 - 8, \tag{4.21}$$

with $\zeta(k) \triangleq \big(s_1(k) + js_2(k)\big)^4$. As a result, an estimation error is committed in the estimate $\hat{\theta}$. The large-sample MSE of this variation can be evaluated by a small-error analysis of the contrast function. The error analysis in the general case is carried out in [3]. Here, we focus on our particular contrast function (4.14). For simplicity, in the rest of this section we write $J = J_{\mathrm{EML}}$. Its sample counterpart is given by

$$\hat{J} = \hat{\gamma} \mathbb{R}\mathrm{e}\big(\hat{\xi}_4(\mathbf{s})\big). \tag{4.22}$$

Also, similarly as in (4.7): $\hat{J}' = 4\hat{\gamma}\mathbb{I}\mathrm{m}\big(\hat{\xi}_4(\mathbf{s})\big)$ and $\hat{J}'' = -16\hat{\gamma}\mathbb{R}\mathrm{e}\big(\hat{\xi}_4(\mathbf{s})\big)$, which are easily arrived at by successive differentiation of (4.20) with respect to $\hat{\theta}$. The estimated angle is actually the solution of the estimating equation $\hat{J}'(\hat{\theta}) = 0$. Approximating this derivative around the true value θ by means of its Taylor series expansion yields:

$$\hat{J}'(\hat{\theta}) \approx \hat{J}'(\theta) + \hat{J}''(\theta)(\hat{\theta} - \theta). \tag{4.23}$$

Therefore: $\hat{J}''(\theta)(\hat{\theta}-\theta) \approx -\hat{J}'(\theta)$. Now, as T tends to infinite, and for i.i.d. components, the CLT [113, 120] guarantees that $\sqrt{T}\hat{J}'(\theta)$ converges in distribution to the normal random variable with zero mean and variance $16\gamma^2 \mathrm{Var}\big[\mathbb{I}\mathrm{m}\big(\zeta\big|_{\hat{\theta}=\theta}\big)\big]$:

$$\hat{J}'(\theta) \xrightarrow[T\to\infty]{\mathrm{d}} \mathrm{N}\left(0, \frac{16\gamma^2}{T}\mathrm{Var}\big[\mathbb{I}\mathrm{m}\big(\zeta\big|_{\hat{\theta}=\theta}\big)\big]\right). \tag{4.24}$$

Since $\mathbb{I}\mathrm{m}\big(\zeta\big|_{\hat{\theta}=\theta}\big) = 4(x_1^3 x_2 - x_1 x_2^3)$ (because when $\hat{\theta} = \theta$, $\mathbf{s} = \mathbf{x}$), and the sources are independent, the asymptotic variance of $\hat{J}'(\theta)$ reduces to $(16\gamma)^2(\mu_{60}^x + \mu_{06}^x - 2\mu_{40}^x\mu_{04}^x)/T$. On the other hand, the SLLN [113, 120] ensures that $\hat{J}''(\theta)$ converges with probability one to its expected value:

$$\hat{J}''(\theta) \xrightarrow[T\to\infty]{\mathrm{wp1}} -16\gamma^2. \tag{4.25}$$

In conclusion, the MSE is given by $\mathrm{E}[(\hat{\theta}-\theta)^2] \approx -\mathrm{E}[\hat{J}'^2(\theta)]/\mathrm{E}[\hat{J}''(\theta)]^2$. With the values calculated above, this reduces to:

$$\mathrm{MSE}_{\mathrm{EML}} \triangleq \mathrm{E}[(\hat{\theta}_{\mathrm{EML}} - \theta)^2] \xrightarrow[T\to\infty]{} \frac{\mu_{60}^x + \mu_{06}^x - 2\mu_{40}^x\mu_{04}^x}{T(\kappa_{40}^x + \kappa_{04}^x)^2}. \tag{4.26}$$

Remark that this value is identical to the variance obtained from the Gaussian approximation to the estimator's asymptotic pdf (Section 3.3.5). Furthermore, this equivalence between large-sample MSE and variance verifies the asymptotic unbiasedness of the estimator.

4.6 Summary and Conclusions

In this chapter, we have interpreted the EML estimator in terms of optimization principles. Two equivalent criteria have been shown to be solved in closed-form by the EML. The first is an output cumulant matching principle, whereby the mismatch between the output kurtosis sum and the source kurtosis sum is minimized. This optimal (in the mean square sense) criterion may be used, in contrast to the general case, in a truly blind environment, for the source statistic of interest does not need to be known in advance, but can be estimated from the available data. The maximization of a contrast function constitutes the second criterion, featuring two remarkable connections. Firstly, the EML contrast function has been proved to expand the applicability of another contrast function defined and studied by Moreau and Comon, which was only valid for sources with the same known kurtosis sign. Simulations have endorsed this interesting outcome. Secondly, the EML contrast function has been shown to correspond to the log-likelihood of the whitened data under the conditions of the AML development, evidencing that the EML is indeed the ML estimator of the unknown parameter under such conditions. Lastly, the asymptotic results obtained in the previous chapter have been confirmed through the small-error analysis of the EML cost function.

Chapter 5

Combined Estimation Strategy

5.1 Introduction

The fourth-order estimators studied in the previous chapters, namely AML, EML, and CF, all share a common drawback: they cannot be applied when the sks is zero. This occurs because the modulus of the centroids on which those estimators are based are all proportional to the sks, rendering the estimation via these procedures meaningless when this source statistic vanishes. To overcome this shortcoming, another 4th-order estimator, the so-called alternative AML (AEML), is put forward (Section 5.2). As we will see, the centroid modulus of the AEML new estimator depends on the source kurtosis difference. The idea is then to devise an estimation strategy (Section 5.3) whereby the EML and the AEML are combined in an attempt to avoid the performance dependency on the cited source statistics. A decision rule is developed to choose between the two estimators. We derive a suboptimal rule, based on a number of experiments, and then check that is indeed close to the asymptotically optimal rule in many cases.

5.2 An Alternative Fourth-Order Estimator

We saw in Section 3.3.3 that the EML centroid can be expanded as a particular complex linear combination of the whitened-vector 4th-order cumulants [eqn. (3.34)]. In the spirit of this observation, we now seek a different complex linear combination of the whitened-signal 4th-order cumulants that supplies us with an explicit expression for the unknown parameter θ. This gives rise to the estimator presented in Section 5.2.1. Section 5.2.3 analyzes its asymptotic performance, and determines the relevant parameters of its asymptotic pdf. The cost function associated to the new estimator is found and studied in Section 5.2.4.

5.2.1 The alternative EML

Let us define the following 4th-order cumulant complex linear combination:

$$\xi_4'(\mathbf{z}) \triangleq (\kappa_{40}^z - \kappa_{04}^z) + 2j(\kappa_{31}^z + \kappa_{13}^z); \tag{5.1}$$

From cumulant equations (3.11), it is easy to realize that:

$$\xi_4'(\mathbf{z}) = \mathrm{e}^{j2\theta}\xi_4'(\mathbf{x}), \tag{5.2}$$

where, according to (5.1),

$$\xi_4'(\mathbf{x}) = \eta \triangleq \kappa_{40}^x - \kappa_{04}^x, \tag{5.3}$$

since the source components are independent. The term η is the *source kurtosis difference (skd)*. Accordingly, the following estimator of θ can be used when the skd is not null:

$$\hat{\theta}_{\mathrm{AEML}} = \frac{1}{2}\arg\big(\xi_4'(\mathbf{z})\big). \tag{5.4}$$

Note that now it is not necessary to estimate source statistic η, since it can only cause a $\pm\pi/2$ radian bias, which does not affect the source separation (Section 3.2.1). Observe also that $\xi_4'(\mathbf{z})$ can be computed in compact form as:

$$\xi_4'(\mathbf{z}) = \mathrm{E}[(z_1^2 + z_2^2)(z_1 + jz_2)^2] = \mathrm{E}[r^4\mathrm{e}^{j2\phi}]. \tag{5.5}$$

By analogy with Definition 4.1, we call (5.1)/(5.5) *alternative fourth-order centroid (AFOC)*. Due to these similarities with the EML, estimator (5.4) is termed *alternative EML (AEML)*.

Links

Estimator (5.4) is not new. An apparently different but actually identical estimator was first proposed independently by Comon in [44] and [45]. This equivalent estimator, that we call *CF'*, is expressed as:

$$|\omega_k| = (-1)^k \frac{|\rho|}{2} + \sqrt{\frac{|\rho|^2}{4} + 1}, \qquad \arg\{\omega_k\} = \arg\{\rho\} + k\pi, \quad k \in \{0, 1\}, \tag{5.6}$$

with

$$\rho = \frac{\kappa_{40}^z - \kappa_{04}^z}{\kappa_{31}^z + \kappa_{13}^z}, \tag{5.7}$$

and $\omega = \mathrm{tg}\,\theta$ (note that the end of expression (5.6) should actually say $\arg\{\omega_k\} = -\arg\{\rho\} + k\pi$, which does not affect the solution in the real case).[1]

The same estimator was independently reached in the work of Clarke and colleagues [118] from the approximate likelihood function derived by Harroy and Lacoume [69]. As we will check in Section 5.2.4, the AEML is (*asymptotically*) the ML estimator of θ when the sources are symmetric, have opposite kurtosis and the 4th-order Gram-Charlier expansion of their jpdf is used.

By contrast, a more clear, rigorous vision is provided herein. First, we have proven above that the applicability of the AEML estimator are not limited to the restrictive conditions mentioned in the previous paragraph; indeed, it is valid as long as the skd is not null. Further, here we endow the estimator with a novel reformulation in terms of 4th-order cumulant complex centroids, coherent with the notation used for the other estimators considered in this work. This allows us: 1) to easily relate estimator (5.4) to the estimators studied in the last chapter, 2) to provide an insightful geometric interpretation (Section 5.2.2) unnoticed in such references, which is analogous to the geometrical explanations of Section 3.3.4 for the EML, 3) to simplify the analysis of its asymptotic performance (Section 5.2.3), which, again, is very similar to that of the EML (Section 3.3.5) and 4) to obtain the optimization principles from which it is derived (Section 5.2.4). In addition, we anticipate that the centroid notation will prove crucial in obtaining a practical decision rule for the combined estimation strategy considered in Section 5.3.

[1] Effectively, there is a mistake is such references: [44, expression (13)] should actually read "$\arg\{\theta_k\} = -\arg\{\rho\} + k\pi$" when the second formula of ρ (eqn. (15) therein, eqn. (5.7) above) is employed. The same comments apply for expression (23) in [45]. Remark that we have borne in mind that in [44, 45] Comon's parameterization of the unitary matrix differs from ours, and also that his definition of the complex cumulants is different from the one we will use in Chapter 8.

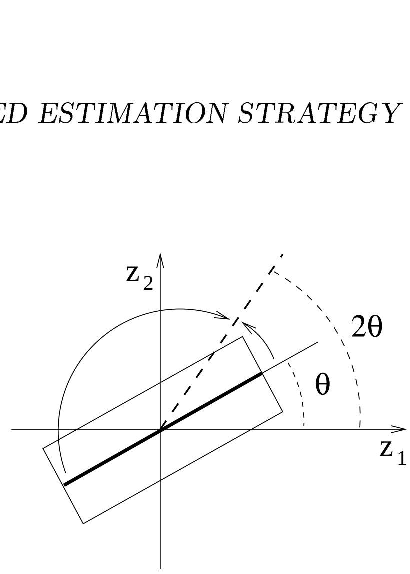


Figure 5.1: AEML centroid orientation (straight dashed line) for two symmetric sources with different kurtosis sign.

5.2.2 Geometrical interpretation

Similarly as for the EML (Section 3.3.4), estimator (5.4) can easily be explained in geometrical terms when the source signals present symmetrical distributions and, additionally in this case, their tail is different. Consider an orthogonal mixture of this type of sources, such as a uniform and a Laplacian distribution. For this situation, the highest density in the source scatter diagram occurs around $\phi' = 0$ and $\phi' = \pi$ (or $\phi' = \pi/2$ and $\phi' = 3\pi/2$, depending on which source is represented at each axis; this is unimportant). In the scatter diagram of the whitened outputs, this corresponds to the lines $\phi = \theta$ and $\phi = \theta + \pi$, as depicted in Figure 5.1. On the other hand, the sample approximation of centroid (5.5) induces over the whitened scatter-diagram points (r_k, ϕ_k), $k = 1, \ldots, T$, the mapping:

$$r_k e^{j\phi_k} \;\mapsto\; r_k^4 e^{j2\phi_k}. \tag{5.8}$$

Consequently, the directions of maximum density are both mapped onto $\phi = 2\theta$. It follows that centroid (5.5) will show the same orientation, half its phase being an estimate of the sought angular parameter.

The above intuitive rationale holds for symmetric sources with different kurtosis sign. This does not mean that the AEML estimator can only be applied in such cases. As we have seen before, the estimator is applicable provided the skd is not null.

5.2.3 Asymptotic analysis

Bias

The AEML estimator is shown to be *unbiased for any sample size* when there is at least one symmetric source. The proof goes along the same lines as that in Section 3.3.5 for the EML estimator.

Basic asymptotic results

Analogously, for i.i.d. observations, the SLLN [113] ensures that

$$\hat{\xi}'_4(\mathbf{z}) \xrightarrow[T\to\infty]{\text{wp1}} \xi'_4(\mathbf{z}) = \eta\, \mathrm{e}^{j2\theta}, \tag{5.9}$$

where $\hat{\xi}'_4(\mathbf{z})$ is the sample estimate of the whitened AFOC:

$$\hat{\xi}'_4(\mathbf{z}) = \frac{1}{T}\sum_{k=1}^{T} r_k^4 \mathrm{e}^{j2\phi_k}, \tag{5.10}$$

and then

$$\hat{\theta}_{\text{AEML}} \xrightarrow[T\to\infty]{\text{wp1}} \theta. \tag{5.11}$$

Hence, as long as $\eta \neq 0$ (i.e., the sources have different kurtosis), the estimator is *strongly consistent* and *asymptotically unbiased.*

Asymptotic pdf

For the AEML estimator (5.4), the sample estimate (5.10) of the corresponding whitened-observation centroid can be decomposed as:

$$\hat{\xi}'_4(\mathbf{z}) = \hat{\xi}'_4(\mathbf{x})\, \mathrm{e}^{j2\theta}, \tag{5.12}$$

where

$$\hat{\xi}'_4(\mathbf{x}) = \frac{1}{T}\sum_{k=1}^{T} r_k^4 \mathrm{e}^{j2\phi'_k} = \frac{1}{T}\sum_{k=1}^{T} \big(x_1^2(k) + x_2^2(k)\big)\big(x_1(k) + jx_2(k)\big)^2. \tag{5.13}$$

Therefore, from (5.4):

$$\Delta\hat{\theta}_{\text{AEML}} \triangleq \hat{\theta}_{\text{AEML}} - \theta = \frac{1}{2}\angle\hat{\xi}'_4(\mathbf{x}). \tag{5.14}$$

From this point on, the analysis is identical to that of the EML, but now:

$$\delta = 2\Delta\hat{\theta}_{\mathrm{AEML}}, \tag{5.15}$$

and parameters $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are redefined according to (5.13):

$$\tilde{\omega}_1 = x_1^4 - x_2^4 \tag{5.16}$$

$$\tilde{\omega}_2 = 2(x_1^3 x_2 + x_1 x_2^3). \tag{5.17}$$

Hence, this time:

$$\begin{aligned} \mathrm{E}[\tilde{\omega}_1] &= \kappa_{40}^x - \kappa_{04}^x = \eta, & \mathrm{E}[\tilde{\omega}_1^2] &= \mu_{80}^x + \mu_{08}^x - 2\mu_{40}^x\mu_{04}^x \\ \mathrm{E}[\tilde{\omega}_2] &= 0, & \mathrm{E}[\tilde{\omega}_2^2] &= 4(\mu_{60}^x + \mu_{06}^x + 2\mu_{40}^x\mu_{04}^x). \end{aligned} \tag{5.18}$$

These values, together with eqns. (3.89) and (3.66), yield the asymptotic pdf of the AEML estimator.

Gaussian approximation (3.67) is also valid for the AEML asymptotic pdf. Now, the new values of the parameters in eqns. (5.18) must be used, yielding the asymptotic variance:

$$\mathrm{Var}[\hat{\theta}_{\mathrm{AEML}}] \xrightarrow[T\to\infty]{} \sigma_{\mathrm{AEML}}^2 = \frac{\mu_{60}^x + \mu_{06}^x + 2\mu_{40}^x\mu_{04}^x}{T(\kappa_{40}^x - \kappa_{04}^x)^2}. \tag{5.19}$$

The small-error analysis on the cost function associated to the AEML estimator, which is explored in Section 5.2.4, endorses this result.

At this stage, it is interesting to remark that the asymptotic variance (3.69) of the EML estimator may be zero for certain source distributions (as occurs, e.g, for binary sources). This perfect estimation, also referred to as deterministic performance, was already noticed in the experiments of Section 3.3.7 (see 'Performance variation with sks'). By contrast, the AEML estimator cannot enjoy deterministic performance, for the numerator of (5.19) can never be null.

Orthogonal invariance

Since the error (5.14) does not depend on the actual value of the true parameter θ (for any sample size), and along the same lines of the EML estimator (end of Section 3.3.5), the AEML can readily be shown to be orthogonal invariant as well. That is, its performance is detached from the specific value of the sought parameter θ.

5.2.4 Optimization criteria

The following lemma, which is a mere formalization of eqns. (5.2) and (5.5), will be useful in deriving the cost functions associated to the AEML estimator.

Lemma 5.1 (alternative-centroid rotation). If $\mathbf{z} = \mathcal{Q}(\theta)\mathbf{x}$, then $\xi_4'(\mathbf{z}) = \mathrm{e}^{j2\theta}\xi_4'(\mathbf{x})$. As an immediate consequence, if, further, $\mathbf{s} = \mathcal{Q}(-\hat{\theta})\mathbf{z}$, then $\xi_4'(\mathbf{s}) = \mathrm{e}^{-j2\hat{\theta}}\xi_4'(\mathbf{z})$. In particular, $|\xi_4'(\mathbf{s})| = |\xi_4'(\mathbf{z})| = |\xi_4'(\mathbf{x})|$, that is, the centroid modulus is not altered by a rotation.

Now we are ready to prove:

Theorem 5.2 (AEML optimization criteria). *Expression* (5.4) *corresponds to the analytical solutions for the local extremization of the functions:*

$$\tilde{J}_{\mathrm{a}4}(\hat{\theta}) = \left|\xi_4'(\mathbf{s}) - \xi_4'(\mathbf{x})\right|^2 = \left|\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}} - \xi_4'(\mathbf{x})\right|^2, \tag{5.20}$$

or

$$\tilde{J}_{\mathrm{b}4}(\hat{\theta}) = \mathbb{R}\mathrm{e}\big(\xi_4'(\mathbf{x})^*\xi_4'(\mathbf{s})\big) = \mathbb{R}\mathrm{e}\big(\xi_4'(\mathbf{x})^*\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}}\big), \tag{5.21}$$

that is,

$$\hat{\theta}_{\text{AEML}} = \arg\operatorname*{ext}_{\hat{\theta}} \tilde{J}_{\mathrm{a}4} = \arg\operatorname*{ext}_{\hat{\theta}} \tilde{J}_{\mathrm{b}4}, \tag{5.22}$$

where "ext" *denotes the extremum (i.e., either maximum or minimum) values of its argument.*

Proof. Application of Lemma 5.1 together with eqn. (5.3) leads to: $\tilde{J}_{\mathrm{a}4}(\hat{\theta}) =$ $= \big(\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}} - \xi_4'(\mathbf{x})\big)\big(\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}} - \xi_4'(\mathbf{x})\big)^* = |\xi_4'(\mathbf{z})|^2 + |\xi_4'(\mathbf{x})|^2 - 2\mathbb{R}\mathrm{e}\big(\xi_4'(\mathbf{x})^*\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}}\big)$ $= 2|\eta|^2 - 2\tilde{J}_{\mathrm{b}4}(\hat{\theta})$. Hence, the maxima of $\tilde{J}_{\mathrm{b}4}$ correspond to the minima of $\tilde{J}_{\mathrm{a}4}$, and vice versa, so that we only need to show the second part of the theorem. The local extrema of (5.21) are given by the first and second order necessary conditions:

$$\frac{\partial \tilde{J}_{\mathrm{b}4}}{\partial \hat{\theta}} = 0, \qquad \text{and} \qquad \frac{\partial^2 \tilde{J}_{\mathrm{b}4}}{\partial \hat{\theta}^2} \neq 0. \tag{5.23}$$

The derivatives of this function are easily calculated by noting again Lemma 5.1. With this in mind:

$$\frac{\partial \tilde{J}_{\mathrm{b}4}}{\partial \hat{\theta}} = 2\mathbb{I}\mathrm{m}\big(\xi_4'(\mathbf{x})^*\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}}\big), \qquad \text{and} \qquad \frac{\partial^2 \tilde{J}_{\mathrm{b}4}}{\partial \hat{\theta}^2} = -4\mathbb{R}\mathrm{e}\big(\xi_4'(\mathbf{x})^*\xi_4'(\mathbf{z})\,\mathrm{e}^{-j2\hat{\theta}}\big). \tag{5.24}$$

The first derivative vanishes if and only if $e^{j[\angle(\xi_4'(\mathbf{x})^* \xi_4'(\mathbf{z}))-2\hat{\theta}]} = \pm 1$, which occurs at the points: $\hat{\theta}_{\text{opt}} = \frac{1}{2}\angle\big(\xi_4'(\mathbf{x})^* \xi_4'(\mathbf{z})\big) + \frac{n\pi}{2}$, $n \in \mathbb{Z}$. This reduces to (5.4) because $\xi_4'(\mathbf{x})$ is real, and then it can only represent a $\pm\pi/2$ rotation which is 'absorbed' by the term $\frac{n\pi}{2}$. For these values of $\hat{\theta}_{\text{opt}}$ the second derivative of $\tilde{J}_{\text{b}4}$ is $\pm 4|\xi_4'(\mathbf{x})|^2 = \pm 4\eta^2 \neq 0$, so they always correspond to local extrema. ■

As seen in the proof above, the actual value of $\xi_4'(\mathbf{x})$ is immaterial in the estimation of θ by means of procedure (5.4). An equivalent function whose local extrema coincide with the AEML estimator (5.4) is then:

$$J_{\text{AEML}} = \mathrm{Re}\big(\xi_4'(\mathbf{s})\big) = \mathrm{Re}\big(\xi_4'(\mathbf{z})\, e^{-j2\hat{\theta}}\big). \tag{5.25}$$

This is the AEML contrast function.

As to the cumulant matching criteria (5.20), the same comments made in Section 4.2.1 are applicable here. Now, the statistic mismatch being optimized is between the skd and the output kurtosis difference.

Link to ML estimation

Asymptotically, the AEML is the ML estimator of the unknown parameter θ under the constraints of the AML development when the source kurtosis sum is null. To prove this claim, we depart, as done for the EML in Section 4.4, from eqns. (3.15) and (3.18). If the source kurtosis have the same magnitude but opposite sign (zero sks) the approximate log-likelihood function transforms into:

$$\log p_{\mathbf{z}}(\mathbf{z} \,|\, \hat{\theta}) = (\kappa_{40}^x - \kappa_{04}^x) \sum_k \frac{(r_k^4 - 6r_k^2)}{4!\,2} \cos 2(\phi_k - \hat{\theta}) + k'(r), \tag{5.26}$$

where the term $k'(r)$ does not depend on $\hat{\theta}$ but only on the moduli r_k. This function can be further expressed as:

$$\log p_{\mathbf{z}}(\mathbf{z} \,|\, \hat{\theta}) = \frac{T}{4!\,2}\,\eta \Big[\mathrm{Re}\big(\hat{\xi}_4'(\mathbf{s})\big) - 6\hat{\xi}_2(\mathbf{s})\Big] + k'(r), \tag{5.27}$$

in which $\hat{\xi}_4'(\mathbf{s})$ represents the sample estimate of the output AFOC $\xi_4'(\mathbf{s})$:

$$\hat{\xi}_4'(\mathbf{s}) = \frac{1}{T} \sum_k r_k^4 e^{j2(\phi_k - \hat{\theta})}, \tag{5.28}$$

and $\hat{\xi}_2(\mathbf{s})$ is the sample estimate of the second-order centroid $\xi_2(\mathbf{s}) = \mathrm{E}[r^2 \mathrm{e}^{j2(\phi-\hat{\theta})}]$:

$$\hat{\xi}_2(\mathbf{s}) = \frac{1}{T} \sum_k r_k^2 \mathrm{e}^{j2(\phi_k - \hat{\theta})}. \tag{5.29}$$

Note that $\xi_2(\mathbf{s}) = \mathrm{E}[(s_1 + js_2)^2] = \mathrm{E}[(s_1^2 - s_2^2) + 2js_1s_2] = (\mu_{20}^s - \mu_{02}^x) + 2j\mu_{11}^s = 0$, owing to the whiteness constraint. Then, for high enough sample size, $\hat{\xi}_2(\mathbf{s}) \approx 0$ (SLLN), so that this term can be discarded in (5.27). Asymptotically, the log-likelihood function thus becomes:

$$\log p_{\mathbf{z}}(\mathbf{z} \,|\, \hat{\theta}) \underset{T\to\infty}{\approx} \frac{T}{4!\,2} \,\eta \mathrm{Re}\big(\hat{\xi}_4'(\mathbf{s})\big) + k'(r), \tag{5.30}$$

whose derivatives with respect to $\hat{\theta}$ are proportional to the sample estimates of the respective derivatives of the AEML cost function (5.25). So the the maximization of (5.30) is equivalent (besides an irrelevant potential $\pm\pi/2$ bias) to the optimization of the sample estimate of the AEML cost function, which results in the AEML estimator (5.4). Hence, the AEML is the (asymptotic) ML estimator of θ when the source jpdf is expanded via its Gram-Charlier approximation truncated at 4th-order and the sources are symmetric with zero kurtosis sum.

The derivation in [118] is vague in that is not mentioned that the relationship unveiled above only holds asymptotically. Like in (3.39), the estimator presented in [118] features an arbitrary radial weighting function $F(r)$, aimed at improving the algorithm robustness against outliers (e.g., in impulsive noise environments). But again, as occurred for (3.39), the choice of arbitrary $F(r)$ is based on empirical geometric reasoning and, besides the case $F(r) = r^4$, no theoretical link with the ML principle is evident.

Small-error analysis

The small-error analysis of contrast function (5.25) is totally analogous to the analysis carried out for the EML contrast function (4.14) in Section 4.5. Now the sample estimate of $\xi_4'(\mathbf{s})$ is

$$\hat{\xi}_4'(\mathbf{s}) = \frac{1}{T} \sum_{k=1}^{T} \big(s_1^2(k) + s_2^2(k)\big)\big(s_1(k) + js_2(k)\big)^2 = \frac{1}{T} \sum_{k=1}^{T} \zeta'(k), \tag{5.31}$$

where $\zeta'(k) \triangleq \big(s_1^2(k)+s_2^2(k)\big)\big(s_1(k)+js_2(k)\big)^2$. For convenience, in the sequel we simply refer to the sample approximation of contrast function J_{AEML} as $\hat{J}$. Its derivatives are worked out as in (5.24), resulting in $\hat{J}' = 2\mathbb{I}\text{m}\big(\hat{\xi}_4'(\mathbf{s})\big)$ and $\hat{J}'' = -4\mathbb{R}\text{e}\big(\hat{\xi}_4'(\mathbf{s})\big)$. The hypothesis of i.i.d. sources makes possible, as in Section 4.5, the application of the CLT and the SLLN, leading to:

$$\hat{J}'(\theta) \xrightarrow[T\to\infty]{\text{d}} \text{N}\left(0,\, \frac{4}{T}\text{Var}\big[\mathbb{I}\text{m}\big(\zeta'|_{\hat{\theta}=\theta}\big)\big]\right) \tag{5.32}$$

$$\hat{J}''(\theta) \xrightarrow[T\to\infty]{\text{wp1}} -4\text{E}\big[\mathbb{R}\text{e}\big(\zeta'|_{\hat{\theta}=\theta}\big)\big]. \tag{5.33}$$

Since $\mathbb{R}\text{e}\big(\zeta'|_{\hat{\theta}=\theta}\big) = x_1^4 - x_2^4$ and $\mathbb{I}\text{m}\big(\zeta'|_{\hat{\theta}=\theta}\big) = 2(x_1^3x_2 + x_1x_2^3)$, the MSE reduces to:

$$\text{MSE}_{\text{AEML}} \triangleq \text{E}[(\hat{\theta}_{\text{AEML}} - \theta)^2] \xrightarrow[T\to\infty]{} \frac{\mu_{60}^x + \mu_{06}^x + 2\mu_{40}^x\mu_{04}^x}{T(\kappa_{40}^x - \kappa_{04}^x)^2}. \tag{5.34}$$

This value coincides with the variance found in Section 5.2.3 [eqn. (5.19)] from the Gaussian approximation to the asymptotic pdf of the AEML estimator.

5.2.5 Computational complexity

In order to evaluate the computational complexity of the AEML estimator, the distinctions of Section 3.3.6 are also made here. Depending on which form of $\xi_4'(\mathbf{z})$, (5.1) or (5.5), is used to compute its sample estimate, we distinguish two cases:

1. Scatter-plot based expressions. Calculation of $\hat{\xi}_4'(\mathbf{z})$ according to (5.5): term $(z_1+jz_2)^2$: 1 complex product = 4 flops per sample; term $(z_1^2+z_2^2)$: no extra products, since it can be computed by using partial results of the previous term; product of the two terms: 2 real products = 2 flops per sample. Overall: $6T$ flops.

2. Cumulants. Calculation of $\hat{\xi}_4'(\mathbf{z})$ according to (5.1): now the cumulant κ_{22}^z is not required, so the computation of the term z_{1122} is avoided, sparing 1 flop/sample. Overall, this takes then about $7T$ flops.

Hence, as opposed to the EML, the more compact complex-centroid form of the AEML estimator is slightly less computationally demanding that the cumulant form. For both forms of calculating their respective centroids, the AEML is less costly than the EML.

5.2.6 Experiments

We divide the AEML estimator's experimental performance evaluation in three parts: variation with the skd η, variation with angle θ and validation of the asymptotic results.

Performance variation with skd

The performance variation with respect to the skd can be measured with the help of the PRBS sources, similarly as done with the sks in Sections 3.3.7 and 3.4.5 to evaluate the EML and CF estimators, respectively. Again one of the source kurtosis is kept constant at $\kappa_{40}^x = -2$, whereas the kurtosis of the second source is varied between $\kappa_{04}^x = -2$ and $\kappa_{04}^x = 14$. Accordingly, the skd is swept between -16 and 0. The results for performance parameter ISR obtained over 100 MC runs are displayed in Figure 5.2. As predicted, the worst performance occurs for zero skd. The best performance is obtained when the kurtosis values are equal in magnitude but opposite in sign (i.e., $\gamma = 0$, giving in this case $\eta = -4$). It was also observed that if the first source is changed to an asymmetric PRBS (hence, with kurtosis greater than -2), the point of best performance still remains around $\gamma = 0$. The fact that the AEML is the ML estimator of θ when the sks is null, for symmetric sources and under the conditions of the Gram-Charlier development (Section 5.2.4) may account for this improved outcome around $\gamma = 0$. The theoretic asymptotic results (represented by the dashed line in Fig. 5.2) are unable to predict the cited performance amelioration; elsewhere they are accurate.

Performance variation with θ

Figure 5.3 shows the performance dependence of the AEML estimates on θ, with a sample size of $T = 5 \cdot 10^3$, for two different source distributions: Rayleigh-Gaussian and exponential–Laplacian. The solid lines represent the empirical MSE averaged over 100 MC runs; the dashed lines plot the theoretic result from expressions (5.19)/(5.34). The AEML orthogonal invariance is confirmed, as well as the accuracy of the asymptotic expressions. Nevertheless, it was noticed that the theoretic prediction was too pessimistic for many source distributions.

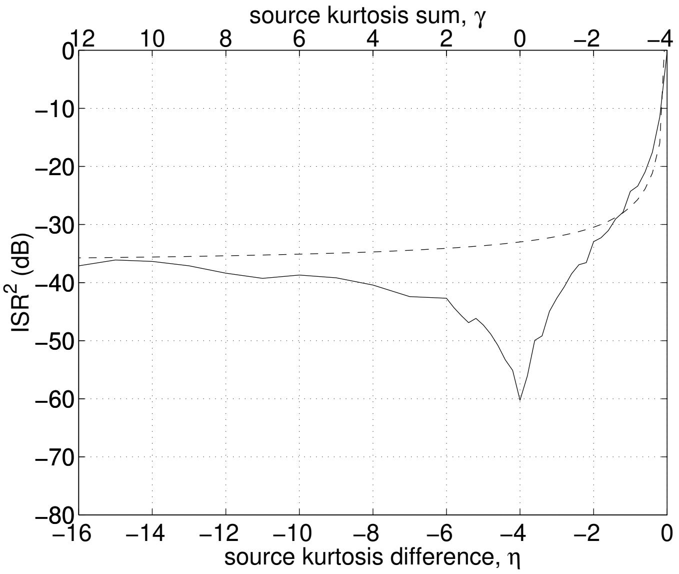


Figure 5.2: Performance of the AEML estimator as a function of the source kurtosis difference, for PRBS sources. $T = 5 \cdot 10^3$ samples, $\theta = 10^o$. Solid line: empirical values of ISR^2 averaged over 100 MC runs per skd value. Dashed line: asymptotic variance/MSE predicted by eqns. (5.19)/(5.34) $[10 \log_{10}(\cdot)]$.

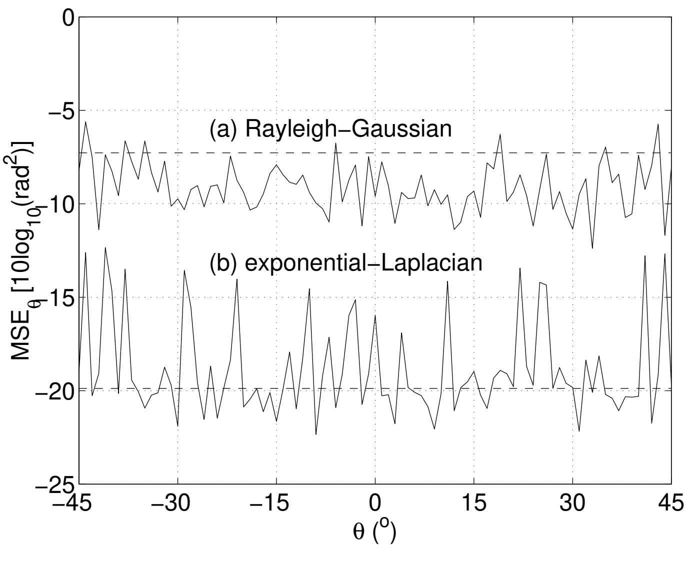


Figure 5.3: Performance of the AEML estimator as a function of the true angle θ, $T = 5 \cdot 10^3$ samples/signal. Solid lines: MSE from the MC simulations, 100 MC runs per angle value. Dashed lines: variance of $\hat{\theta}$ predicted from (5.19)/(5.34) $[10 \log_{10}(\cdot)]$. Source distributions: (a) Rayleigh–Gaussian , (b) exponential–Laplacian.

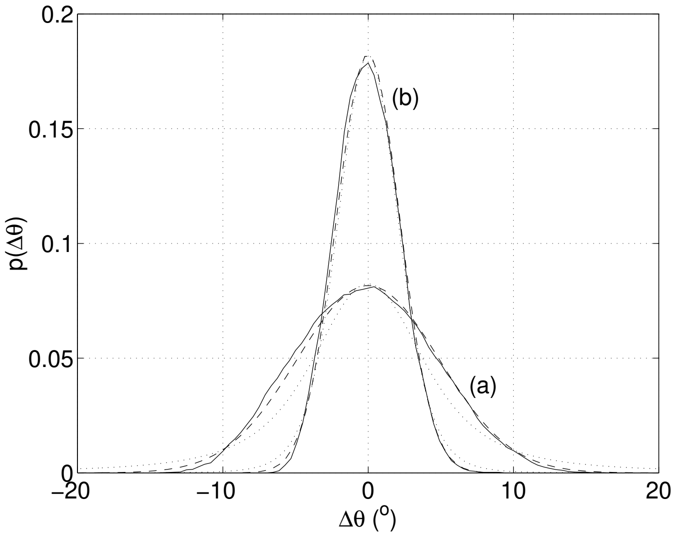


Figure 5.4: Pdf of the AEML error for exponential and asymmetric triangular source distributions, 1000 signal realizations. Solid lines: naive density estimate. Dotted lines: analytic asymptotic pdf (3.66) with parameters (5.18). Dashed lines: Gaussian approximation (3.67) with parameters (5.18). (a) $T = 10^3$ samples/signal, naive pdf estimate window width $h = 11.25^{\mathrm{o}}$; (b) $T = 5 \cdot 10^3$ samples/signal, naive pdf estimate window width $h = 4.5^{\mathrm{o}}$.

Validation of asymptotic pdf

An example to illustrate the validity of the AEML-estimator asymptotic pdf is graphically summarized in Figure 5.4. The same experiment as that of Fig. 3.8 in Section 3.3.7 is carried out, but using the AEML estimator (5.4). The source distributions are now chosen as exponential and asymmetric triangular, and $\theta = 15^{\mathrm{o}}$. The pdf estimation results from 1000 angle estimates and for two different sample sizes are displayed in Fig. 5.4. The precision of the asymptotic pdf approximations is remarkable, and improves as the sample size increases.

5.3 Combined Estimation Strategy

The EML estimator presented in Chapter 3 shows a dependency with the sks γ, so that a performance degradation is observed as this parameter tends to zero. This outcome motivated the development of the alternative 4th-order estimator AEML, carried out earlier in this chapter. Nonetheless, the AEML performance varies with the skd η, and so this estimator also suffers from a comparable deficiency, but with respect to another source statistic. In an attempt to overcome these undesirable dependencies, we now consider the idea of combining both estimation expressions (Section 5.3.1). The choice between the two estimators is to be made through a suitable decision rule, which is analyzed in detail in Sections 5.3.2 and 5.3.3. Some assessment and comparison experiments are reported in Section 5.3.4.

5.3.1 Combining estimators

The idea of combining estimators appeared for the first time in [44,45]. Essentially, the CF and the AEML (in the form of eqns. (5.6)–(5.7), which we jointly refer to as CF') were combined in such reference via the decision rule:

$$|\kappa_{22}^z| \underset{\text{CF'}}{\overset{\text{CF}}{\gtrless}} |\kappa_{31}^z + \kappa_{13}^z|. \tag{5.35}$$

We name this the *combined CF (combCF)* estimator. However, there was not rationale behind this rule other than the conditioning of the denominators of ρ in (3.74) and (5.7).

By contrast, here we propose to combine the EML and the AEML via a more rational scheme. Such combination can be made through a generic rule:

$$f(\mathbf{z}) \underset{\text{AEML}}{\overset{\text{EML}}{\lessgtr}} g(\mathbf{z}), \tag{5.36}$$

in which $f(\mathbf{z})$ and $g(\mathbf{z})$ are certain functions, to be specified, of the whitened vector (data available). The above equation reads: "if f is smaller than g, then estimate the unknown angle through the EML; otherwise, use the AEML." Hence, given a set of whitened signals, these functions determine which estimator should be used.

Obviously, one would like to obtain always the most accurate estimation, so a sensible optimal choice would be to take in each case the estimator which would provide the smallest MSE. However, this depends in both cases on the source statistics (see the asymptotic analyses of Sections 3.3.5 and 5.2.3), and in a blind problem these are unknown. That is, if the source distribution is not known a priori, no functions f and g of the given observation vector can be devised so that they represent the MSE that would be obtained by each estimator. Therefore, we are left with adopting a suboptimal decision strategy based on the available information (Section 5.3.3). But first (Section 5.3.2), we explore the ideal optimal rule, in the sense of the minimization of the MSE, when knowledge about the source distribution is available. In the simulations of Section 5.3.4, the combined estimator put forward next will be compared to the combCF.

5.3.2 Optimal decision rule

The optimal, in the MMSE sense, decision rule corresponds to the criterion:

$$\mathrm{MSE}_{\mathrm{EML}} \underset{\mathrm{AEML}}{\overset{\mathrm{EML}}{\lessgtr}} \mathrm{MSE}_{\mathrm{AEML}}. \tag{5.37}$$

The (asymptotic) MSE's where obtained in eqns. (4.26) and (5.34), respectively, and coincide with the asymptotic variances found in eqns. (3.69) and (5.19). This is hence an asymptotically-optimal MMSE decision rule, which, in the nomenclature of the previous section, can be associated with respective functions f_{opt} and g_{opt}. If the source distributions are known in advance, rule (5.37) tells which estimator will provide MMSE performance.

Though most appealing, this rule does not have a practical interest in a really blind situation. The reason is that the expected MSE values cannot be estimated from the observed data. As an illustration of this fact, let us compute the value of the skd η, which appears at the denominator of the AEML MSE, with respect to the whitened observations. As in (3.25), the skd (5.3) can be written as

$$\begin{aligned}
\xi_4'(\mathbf{x}) \underset{\substack{\uparrow \\ (5.5)\ \&\ (3.12)}}{=} & \mathrm{E}[r^4 \mathrm{e}^{j2\phi'}] = \mathrm{E}[(x_1^4 - x_2^4) + j2(x_1^3 x_1 + x_1 x_2^3)] \underset{\substack{\uparrow \\ \kappa_{31}^x = \kappa_{13}^x = 0}}{=} \mathrm{E}[(x_1^2 + x_2^2)(x_1^2 - x_2^2)] \\
&= \mathrm{E}[r^2 \mathrm{Re}(r^2 \mathrm{e}^{j2\phi'})] \underset{\substack{\uparrow \\ (3.13)}}{=} \mathrm{E}[r^2 \mathrm{Re}(r^2 \mathrm{e}^{j2\phi} \mathrm{e}^{-j2\theta})] \\
&\underset{\substack{\uparrow \\ (3.12)}}{=} \mathrm{E}\Big[(z_1^2 + z_2^2)\big[(z_1^2 - z_2^2)\cos 2\theta + 2z_1 z_2 \sin 2\theta\big]\Big].
\end{aligned} \tag{5.38}$$

Consequently, as opposed to γ, parameter η cannot be determined only from the whitened observations, since it depends on the actual angle as well, which is precisely what we want to estimate. The numerators of eqns. (4.26) and (5.34) also show a similar dependency on the unknown parameter when expanded as a function of $\mathbf{z}$. Hence, rule (5.37) cannot be expressed as a function of $\mathbf{z}$ being, at the same time, free of θ. That is, $\mathrm{MSE}_{\mathrm{EML}} = f_{\mathrm{opt}}(\mathbf{z}, \theta)$ and $\mathrm{MSE}_{\mathrm{AEML}} = g_{\mathrm{opt}}(\mathbf{z}, \theta)$. In conclusion, a suboptimal practical principle is necessary. Optimality is to be swapped for practical applicability.

5.3.3 A suboptimal decision rule

An empirically-based suboptimal decision rule with PRBS sources

In a bid to obtain a tractable practical guideline to decide between the 4th-order estimators EML and AEML, the following experiment is conducted. The performance of such estimators is to be evaluated for different values of sks γ and skd η, which can be achieved by resorting to the ideas of [135] (Appendix B), as done in Sections 3.3.7, 3.4.5 and 5.2.6. One of the sources is chosen as a binary sequence of equiprobable symbols, and hence $\kappa_{40}^x = -2$. The other source is also a PRBS, but with varying kurtosis between -2 and 14. This gives $\gamma \in [-4, 12]$ and $\eta \in [-16, 0]$. Each signal realization is composed of $T = 5 \cdot 10^3$ samples. At each kurtosis value, separations are performed by the two methods on the same orthogonal mixtures of $\theta = 30^{\mathrm{o}}$ over 100 independent MC runs.[2] Following each separation, the ISR is computed from the results obtained by each method, and then averaged after the 100 realizations.

Figure 5.5 shows the results. As appreciated before, the EML performance worsens around $\gamma = 0$ (Section 3.3), and analogously for the AEML about $\eta = 0$ (Section 5.2). Taking into account that:

$$|\xi_4(\mathbf{z})| = |\kappa_{04}^x + \kappa_{04}^x| = |\gamma| \tag{5.39a}$$

$$|\xi_4'(\mathbf{z})| = |\kappa_{04}^x - \kappa_{04}^x| = |\eta|, \tag{5.39b}$$

the performance of both methods decreases when their respective centroid locations are close to the origin of the complex plane. For $\gamma < -2$, the EML shows better performance than the AEML, and conversely when $\gamma > -2$. So it seems there is an alternation in best performance when the signs of source kurtosis pass from being equal (EML is better) to being different (AEML is better). That is, based on these results

[2] Once more the actual value of θ is unimportant, since both the EML and the AEML have been proven to be orthogonal invariant (Sections 3.3.5 and 5.2.3).

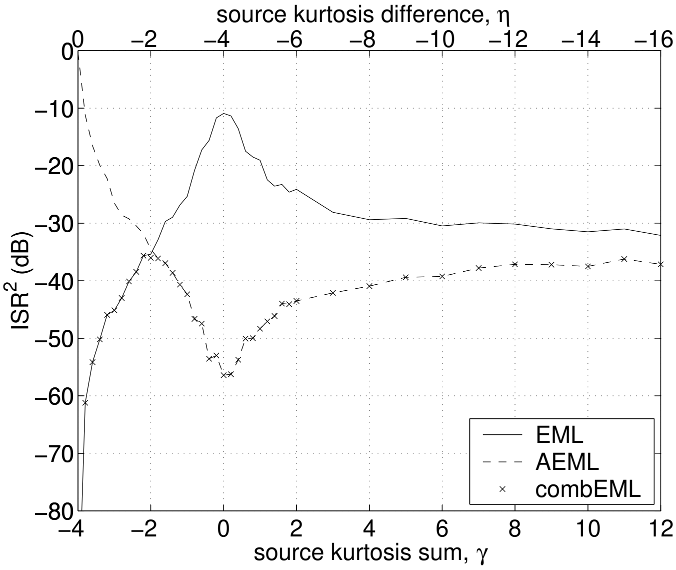


Figure 5.5: Performance of EML, AEML and combined estimators vs. sks.

one could derive the following decision rule between the two methods:

$$\kappa_{40}^{x}\,\kappa_{04}^{x} \underset{\text{AEML}}{\overset{\text{EML}}{\gtrless}} 0. \tag{5.40}$$

Some straightforward algebraic manipulations on the above expression lead to:

$$(\kappa_{40}^{x} + \kappa_{04}^{x})^2 \underset{\text{AEML}}{\overset{\text{EML}}{\gtrless}} (\kappa_{40}^{x} - \kappa_{04}^{x})^2. \tag{5.41}$$

Bearing in mind eqns. (5.39), this is turned into the following practical decision rule, i.e., a principle that can be applied from the available information:

$$|\xi_4(\mathbf{z})| \underset{\text{AEML}}{\overset{\text{EML}}{\gtrless}} |\xi_4'(\mathbf{z})|. \tag{5.42}$$

Thus, if the modulus of the EML centroid (3.21) is bigger than the modulus of the AEML centroid (5.5), then the former should be applied, and conversely. Intuitively, this empirical rule of thumb makes sense: choose the estimator with the largest centroid modulus. We refer to the joint estimation procedure associated with rule (5.42) as the *combined EML (combEML)* estimator. Fig. 5.5 also shows the result obtained with this combined strategy under the same conditions quoted above. It consistently keeps the best performance over all range of sks. Additional theoretic considerations and further experiments on discrete as well as continuous distributions corroborate the decision rule (5.42), as we discuss next.

Validation on the AML principle

The AML function (3.18) may be expressed as:

$$\log p_{\mathbf{z}}(\mathbf{z}(k) \,|\, \hat{\theta}) = k(r) + \gamma A(r) \cos 4(\phi_k - \hat{\theta}) + \eta B(r) \cos 2(\phi_k - \hat{\theta}), \tag{5.43}$$

in which $k(r)$, $A(r)$ and $B(r)$ denote functions of the moduli r_k. As we saw in Section 4.4, the second term on the right-hand side of equation (5.43) corresponds to the EML estimator when $\eta = 0$. Accordingly (Section 5.2.4), the third term results, asymptotically, in the AEML estimator when $\gamma = 0$. Hence, in the light of expression (5.43) the relative value of γ and η indicates the significance of the EML and the AEML solutions in the AML criterion. On the other hand, and due to the properties of ML estimation [111], it would be reasonable to expect a better performance from that estimator closest to the ML criterion. This 'closeness', as hinted above, can be measured by comparing the values of γ and η, which is thus in agreement with the rule derived in the previous section.

Very similar conclusions are reached [72] from the Edgeworth expansion [48,78,126] of the mutual information [34,48,55] of the separator output.

Validation on PRBS sources

What rule (5.42) essentially says [via its equivalent expression (5.40)] is that the EML should be used when the sources have the same kurtosis sign (same tail type), and the AEML otherwise. We now intend to see to what extent this simplified rule is faithful to the asymptotically optimal rule (5.37) when PRBS sources are processed.

The normalized statistics in Appendix A.4 allows the computation of the optimal decision rule (5.37) for this kind of sources. The solid curves in Figure 5.6 are the loci of the PRBS-source kurtosis for which both sides of (5.37) are equal.[3] That is, the curves correspond to the limits between the source-kurtosis regions where the EML and the AEML should be applied. Similarly, the dashed line represents the boundary of practical rule (5.42). Since the change takes place when the sign of sources kurtosis become distinct [eqn. (5.40)], this boundary is always set at the lines delimited by $\kappa_{40}^x = 0$ and $\kappa_{04}^x = 0$.

[3] Note that in order to generate the plot of Fig. 5.5 the kurtosis of the first source was kept constant at $\kappa_{40}^x = -2$.

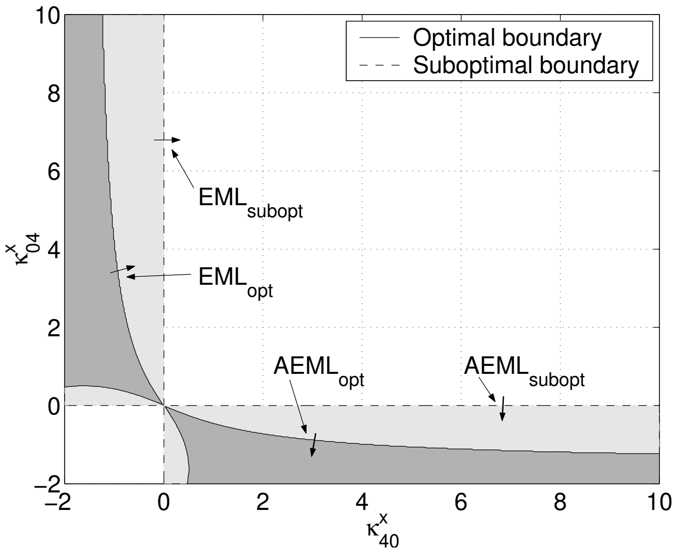


Figure 5.6: Boundaries between the EML and the AEML application regions on the source kurtosis map for PRBS sources. Solid line: optimal MMSE rule (5.37). Dashed line: suboptimal practical rule (5.42).

Observe how the trend shown by the optimal boundary remains very close to the limit set by the practical principle, indicating that rule (5.42) is indeed a fair approximation of the MMSE asymptotically optimal rule (5.37) for PRBS sources.

Validation on continuous-distribution sources

Up to now, the proposed decision rule has only been tested on PRBS, which are by nature discrete distributions. Now, we test it on sources with continuous distributions (the binary distribution is also included for the sake of completeness). We use the well-known distributions whose statistics are summarized in Appendix A.4. From those statistics, the values of $T\sigma_{\mathrm{EML}}$ and $T\sigma_{\mathrm{AEML}}$ analytically given by eqns. (3.69) and (5.19) are computed for the 36 different source combinations (the source ordering is not important). The results appear in Table 5.1. The boldface elements denote signal combinations for which the choices made via optimal rule (5.37) and via suboptimal principle (5.42) do not agree. Only 3 of such cases occur,[4] and all of them involve a Rayleigh source together with a short-tailed source (binary, sinusoid and uniform). This

[4]Combinations with the same distributions but in different order are not considered.

outcome could be due to the fact that the Rayleigh distribution is almost mesokurtic (its kurtosis is very close to zero). On the other hand, the practical rule (5.42) says that either estimator can be used when there is a Gaussian source (this was confirmed in Fig. 5.6 with the PRBS approximation to signals of zero kurtosis). By contrast, Table 5.1 now shows that the best estimator when one Gaussian source is present is indeed the EML, at least for the other source distributions tested here.

5.3.4 Simulations

The experiments reported next contrast the combEML with the combCF and with two other methods not based on closed-form estimation, namely HOEVD and JADE.

Comparison combEML–combCF

The following simulation compares the combined estimation procedures combEML and combCF. The combEML is composed of estimators EML (3.32) and AEML (5.4) plus decision rule (5.42). On the other hand, the combCF consists of estimators CF [(3.73) with (3.74)] and CF' [(3.73) with (5.7)], combined through principle (5.35). PRBS sources are used to obtain a range of values of sks and skd. We select $\theta = 10^{o}$. The sample size is fixed at $T = 5 \cdot 10^3$ and 200 MC iterations are run. In the first example the first source kurtosis is chosen as $\kappa_{04}^{x} = -2$, whereas in the second example $\kappa_{40}^{x} = 1$. In both cases the second source kurtosis is swept between -2 and 12. The average values of performance parameter ISR are plotted in Figures 5.7 and 5.8, respectively.

In Fig. 5.7 ($\kappa_{40}^{x} = -2$) the combEML behaves as in the experiment of Section 5.3.3 (Fig. 5.5). It maintains the best performance out of the two estimators it is based upon over the whole sks range. For the part of the combCF, matters are different. This combined estimator is unable to track the optimum performance for all sks values. It fails while the two sources have the same kurtosis sign (sks below -2). Note that, as claimed in Section 5.2.1, the AEML and the CF' are identical.

In Fig. 5.8 ($\kappa_{40}^{x} = 1$), the combCF performance improves relative to the previous case. Problems are observed in the region around $\gamma = 2$, i.e., when both sources have the same kurtosis value. By contrast, the combEML shows again a consistent lowest-ISR performance versus sks, evidencing, in turn, its robustness with respect

		$\kappa_{40}^x < 0$				$\kappa_{40}^x = 0$	$\kappa_{40}^x > 0$		
		Bin	Sin	Uni	Tri	Gau	Ray	Lap	Exp
$\kappa_{04}^x < 0$	Bin	0.000	0.041	0.123	0.729	2.500	**5.293**	79.000	15.500
		∞	26.000	13.214	7.411	5.500	**5.809**	4.120	4.438
	Sin	0.041	0.056	0.131	0.913	3.778	**9.243**	33.111	11.877
		26.000	∞	130.635	22.749	11.778	**11.173**	5.457	5.236
	Uni	0.123	0.131	0.214	1.217	5.595	**15.316**	22.302	10.263
		13.214	130.635	∞	58.955	20.595	**17.876**	6.545	5.811
	Tri	0.729	0.913	1.217	4.120	25.907	118.624	12.140	7.906
		7.411	22.749	58.955	∞	105.907	64.540	9.840	7.276
$\kappa_{04}^x = 0$	Gau	2.500	3.778	5.595	25.907	∞	288.319	7.667	6.278
		5.500	11.778	20.595	105.907	∞	936.540	15.667	9.278
$\kappa_{04}^x > 0$	Ray	**5.293**	**9.243**	**15.316**	118.624	288.319	93.721	6.918	5.856
		5.809	**11.173**	**17.876**	64.540	936.540	∞	19.861	10.423
	Lap	79.000	33.111	22.302	12.140	7.667	6.918	3.000	3.049
		4.120	5.457	6.545	9.840	15.667	19.861	∞	51.444
	Exp	15.500	11.877	10.263	7.906	6.278	5.856	3.049	2.556
		4.438	5.236	5.811	7.276	9.278	10.423	51.444	∞

Table 5.1: Asymptotic values of $T\sigma^2_{\text{EML}}$ (top figures) and $T\sigma^2_{\text{AEML}}$ (bottom figures), in squared radians, for various source distributions. Values are computed via analytical expressions (3.69) and (5.19), and the true statistics of Tables A.1 and A.2. Figures in boldface indicate source combinations for which the decision of simplified rule (5.42) would differ from that of asymptotically optimal rule (5.37).

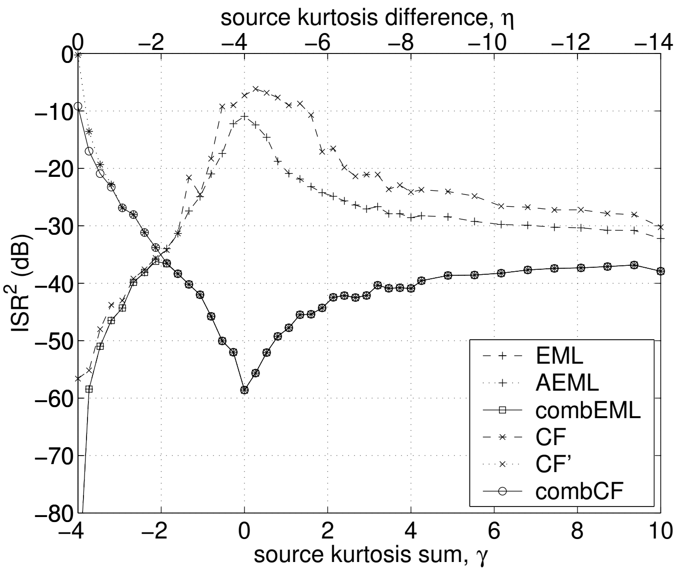


Figure 5.7: Performance of the combEML and the combCF vs. sks, for $\kappa_{40}^{x} = -2$. PRBS sources, $\theta = 10^{o}$, $T = 5 \cdot 10^{3}$ samples/signal, 200 MC runs.

to the specific values of the source kurtosis. In this second example, the combEML performance improves that of the combCF for $\gamma > 1$, that is, when both sources have the same kurtosis sign. This outcome is related to the CF exhibiting worse performance than the EML for low angle values (Section 3.4.3).

Comparison with HOEVD and JADE

Next, the combEML is compared with two other methods not based on closed-form estimation, but rather on contrast-function optimization: the HOEVD [48] and JADE [23]. Refer to Sections 2.5.4 and 2.5.5, respectively, for an explanation of both methods.[5]

Continuing with the PRBS-based sks simulations, the first comparison is made in this context. By means of PRBS sources, sks values are simulated between -4 and 10, where the first source is fixed at $\kappa_{40}^{x} = -2$. The sample size is $T = 5 \cdot 10^{3}$, the number

[5]Original MATLAB code implementing HOEVD and JADE was downloaded from the URL's:

`http://www.i3s.unice.fr/~comon/matlab.html`

`http://www-sig.enst.fr/~cardoso/guidesepsou.html`,

respectively. Since only unitary mixtures were considered in these simulations, the pre-whitening part was skipped in both original algorithms. Additionally, the cumulant computation was removed from inside the respective functions, and the cumulants were entered externally as input parameters. The aim of this simple modification was to make all tested methods work from the same cumulant values, so that subsequent comparisons were more significant.

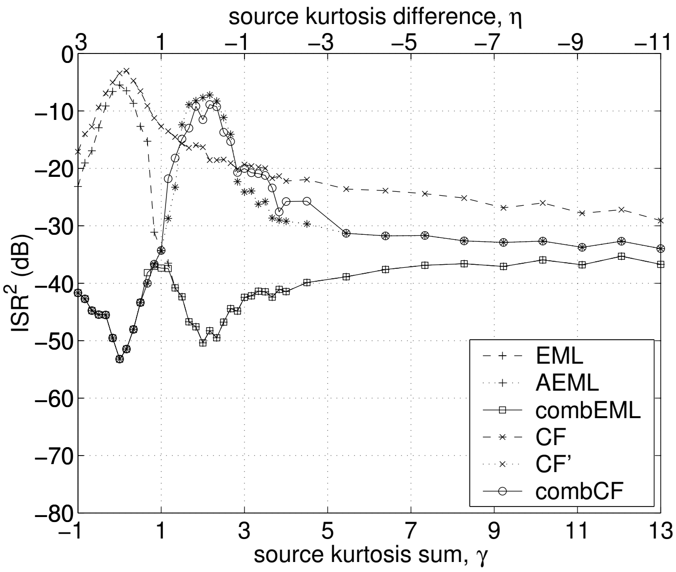


Figure 5.8: Performance of the combEML and the combCF vs. sks, for $\kappa_{40}^{x} = 1$. PRBS sources, $\theta = 10^{\text{o}}$, $T = 5 \cdot 10^3$ samples/signal, 200 MC runs.

of MC iterations is 200, and $\theta = 10^{\text{o}}$. Results appear in Figure 5.9. Interestingly, the combEML outperforms the HOEVD and JADE by about 4 dB (the mean difference of the curves shown is 4.4 dB). The performance of the latter two methods is virtually identical. Remark the equal trends exhibited by all three methods' curves, as well as their deterministic performance for equiprobable-symbol PRBS sources ($\gamma = -4$). Note that, contrary to what is hinted in [34, Section VII-C], deterministic performance is obtained without taking advantage of discreteness.

Similar outcomes are obtained if the first source is fixed at $\kappa_{40}^{x} = 1$, as shown in Figure 5.10. Again, the performances of HOEVD and JADE are identical, and the combEML improves those two methods by about 4 dB (the mean difference between the curves was this time 3.7 dB).

In a final comparison experiment, the values of $T \times \mathrm{E}[(\hat{\theta} - \theta)^2]$ are computed from 500 independent MC runs over signal realizations composed of $T = 10^4$ samples, with a true angle of $\theta = 10^{\text{o}}$. The results are shown in Table 5.2. Boldface figures indicate the source combinations for which the combEML outperforms the other two methods. This occurs in 20 out of the 36 different signal distributions (55.56%), that is, in the

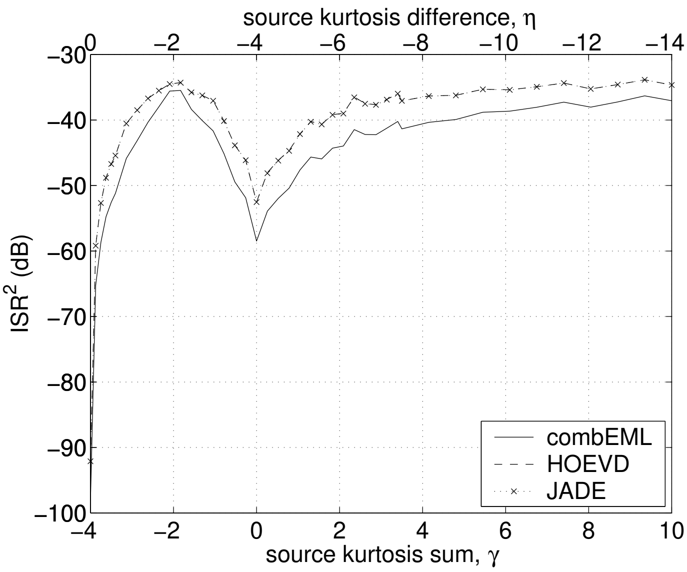


Figure 5.9: Performance of the combEML, HOEVD and JADE vs. sks, for $\kappa_{40}^{x} = -2$. PRBS sources, $\theta = 10^{o}$, $T = 5 \cdot 10^{3}$ samples/signal, 200 MC runs.

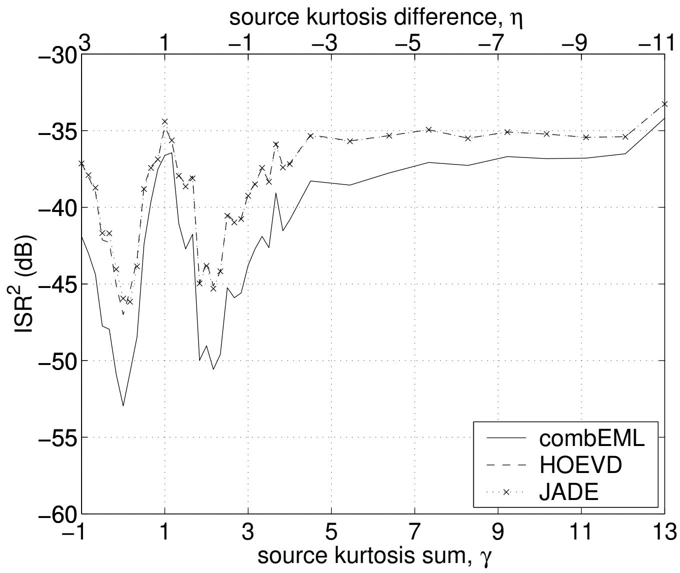


Figure 5.10: Performance of the combEML, HOEVD and JADE vs. sks, for $\kappa_{40}^{x} = 1$. PRBS sources, $\theta = 10^{o}$, $T = 5 \cdot 10^{3}$ samples/signal, 200 MC runs.

majority of cases tested.[6] In turn, these signal combinations can be further classified into:

$$\text{a) Source kurtosis sign:}^{7} \begin{cases} \text{equal:} & 10/16\ (62.50\%) \\ \text{different:} & 8/12\ (66.67\%). \end{cases}$$

$$\text{b) Source kurtosis sum:}^{8} \begin{cases} \text{positive:} & 12/17\ (70.59\%) \\ \text{negative:} & 7/18\ (38.89\%). \end{cases}$$

Therefore, the combEML seems to work better than the HOEVD and JADE for source combinations with positive kurtosis sum, regardless of the specific signs of source kurtosis.

Finally, we can compare the experimental and the theoretical figures of Tables 5.1 and 5.2. Those combEML values of Table 5.2 corresponding to the sources for which the EML is mostly used (essentially those cases for which $\text{sign}(\kappa^x_{40}) = \text{sign}(\kappa^x_{04})$, such as Bin–Sin, Uni–Tri, Lap–Exp, etc., plus others such as Sin–Gau, Uni–Gau, etc.), are very close to the EML analytic asymptotic results of Table 5.1. In the cases that $\text{sign}(\kappa^x_{40}) \neq \text{sign}(\kappa^x_{04})$ (such as Exp–Bin, Exp–Uni, etc.), where the AEML is preferred in the combEML scheme, the results are generally lower than the corresponding AEML figures of Table 5.1. This indicates that in many cases the AEML can actually do better than theoretically predicted, as already noticed in the experiments of Section 5.2.6 (see 'Performance variation with θ'). Hence, the AEML asymptotic analysis may need some sort of refinement in order to account for these discrepancies.

5.4 Summary and Conclusions

In the first part of this chapter, another 4th-order cumulant based closed-form estimator has been put forward, in a bid to overcome the performance degradation of the EML for sks near zero. This estimator, the AEML, had previously been proposed in the literature under a different form. However, here it has been reformulated in terms of complex centroids and other devices used in this work for the analysis of closed-form BSS estimators. This has allowed a deeper, more rigorous understanding of its

[6] Combinations with the same distributions but appearing in different order are not counted in these and the following statistics.
[7] Combinations comprising a Gaussian source are excluded.
[8] The Gaussian–Gaussian combination is not considered.

		$\kappa_{40}^x < 0$				$\kappa_{40}^x = 0$	$\kappa_{40}^x > 0$		
		Bin	Sin	Uni	Tri	Gau	Ray	Lap	Exp
$\kappa_{04}^x < 0$	Bin	**0.000**	**0.040**	**0.121**	**0.679**	2.299	2.777	**1.915**	**3.026**
		0.000	0.089	0.257	0.804	0.974	0.940	2.624	4.601
		0.000	0.089	0.257	0.804	0.974	0.938	2.614	4.591
	Sin	**0.040**	**0.055**	**0.132**	0.869	3.560	5.644	**2.549**	**3.576**
		0.089	0.055	0.146	0.660	1.128	1.293	3.783	5.175
		0.089	0.055	0.146	0.659	1.128	1.289	3.768	5.167
	Uni	**0.121**	**0.132**	0.215	1.241	5.245	8.440	**3.105**	**3.888**
		0.257	0.146	0.215	0.934	1.280	2.046	4.755	5.439
		0.257	0.146	0.215	0.933	1.278	2.061	4.769	5.439
	Tri	**0.679**	0.869	1.241	**3.996**	22.509	19.539	**5.127**	**4.633**
		0.804	0.660	0.934	4.015	9.234	9.293	7.031	5.707
		0.804	0.659	0.933	4.016	9.272	9.324	7.042	5.706
$\kappa_{04}^x = 0$	Gau	2.299	3.560	5.245	22.509	**2501.994**	351.577	6.902	**6.542**
		0.974	1.128	1.280	9.234	2531.774	283.766	6.218	6.598
		0.974	1.128	1.278	9.272	2540.803	293.209	6.219	6.599
$\kappa_{04}^x > 0$	Ray	2.777	5.644	8.440	19.539	351.577	234.390	7.731	6.272
		0.940	1.293	2.046	9.293	283.766	200.965	7.386	5.959
		0.938	1.289	2.061	9.324	293.209	199.179	7.375	5.957
	Lap	**1.915**	**2.549**	**3.105**	**5.127**	6.902	7.731	**3.176**	**2.927**
		2.624	3.783	4.755	7.031	6.218	7.386	3.294	3.650
		2.614	3.768	4.769	7.042	6.219	7.375	3.292	3.642
	Exp	**3.026**	**3.576**	**3.888**	**4.633**	**6.542**	6.272	**2.927**	**3.057**
		4.601	5.175	5.439	5.707	6.598	5.959	3.650	3.305
		4.591	5.167	5.439	5.706	6.599	5.957	3.642	3.299

Table 5.2: Empirical values of $T \cdot \text{MSE}_{\hat{\theta}}$ by the combEML, HOEVD and JADE methods for various source distributions. MSE values are obtained by averaging angle estimates (expressed in radians) over 500 independent MC iterations. Signal realizations are composed of $T = 10^4$ samples. $\theta = 10^{\circ}$. Each 3-element entry corresponds, from top to bottom, to the combEML, HOEVD and JADE. Boldface figures indicate cases where the combEML outperforms the other two methods.

properties and its bonds with other procedures. Benefiting from this common framework, we have analyzed its asymptotic performance, derived the optimization criteria it implicitly stems from, and understood the existing statistical principles behind it.

The second part of the chapter has been devoted to studying the idea of combining the EML and the AEML estimators. Both show a performance variation with respect to certain source statistics. The combination of the estimators according to a statistically sound decision rule has been aimed at eliminating that dependency. Relying on empirical observations, a practical decision rule has been proposed, and later validated by showing to be very close to the optimal, in the MMSE sense, rule in numerous cases. Simulations have demonstrated the superior performance of the resulting combined estimator (combEML) relative to another previously proposed (combCF). As another indication of its performance quality, the combEML also improves contrast-based methods such as HOEVD and JADE in a good number of source distribution combinations.

Chapter 6

General BSS Scenario: Application to the Fetal ECG Extraction Problem

6.1 Introduction

The closed-form estimators studied in the preceding chapters are designed to achieve BSS in the two-signal case. In this chapter we demonstrate that a particular extension, based on iterations over the signal pairs, enables such procedures to tackle the more general BSS scenario where more than two mixtures of (possibly) more than two sources are observed. With the twofold purpose of evaluating their performance and illustrating the use of BSS techniques in a real-life application, we intend to apply the resulting methods to a biomedical engineering problem: the extraction of the fetal electrocardiogram (FECG) signals from cutaneous electrodes placed on the mother's body. A number of BSS techniques are to be tested on real ECG data recorded from a pregnant woman, and compared within this interesting application.

In Section 6.2, the FECG extraction problem is explained, and it is reviewed how this biomedical application accepts a BSS model. The real recordings used in the experiments are presented in Section 6.3. Section 6.4 summarizes the methods employed. We consider the pairwise extension of the closed-form methods developed in the previous chapters and two other well-established blind separation procedures: JADE and HO-EVD. In order to compare in this context this type of methods with other techniques not based on blind signal separation, a classical approach is also considered: Widrow's multi-reference adaptive noise cancellation. The results are summarized in Section 6.5, and discussed in Section 6.6.

6.2 The Fetal ECG Extraction Problem. BSS Model

During pregnancy, monitoring the fetus' heart condition in order to test their well-being and diagnose possible diseases is of paramount importance. An early diagnosis before delivery increases the effectiveness of any subsequent treatment. This requirement calls for non-invasive techniques, among which the fetal electrocardiogram (FECG) is acknowledged by the medical community as a reliable diagnosis tool. The extraction of the *antepartum* FECG can be carried out through electrodes attached to the mother's skin. Unfortunately, the electrode recordings are not only composed of the desired FECG components, but also of a number of other disturbances. The most important are the maternal ECG (MECG) contributions, of considerably higher amplitude than the fetal components. Mother's respiration and electromyographic (EMG) signals (e.g., owing to an uncomfortable position of the patient, uterus contraction, etc.) act as a second source of biological interference. Non-biological interference sources, such as mains coupling and thermal noise from the electronic equipment, corrupt the cutaneous recordings as well. Consequently, the output of the electrodes show the fetal heartbeat contributions buried in an additive mixture of interference, which hinders the observation of the FECG and hence the diagnosis procedure. Appropriate signal processing techniques are thus required in order to recover the wanted FECG components from the corrupted potential recordings.

A variety of different approaches have been proposed to address this problem. Among these classical procedures were techniques such as coherent averaging, matched filtering, auto- and cross-correlation based methods, adaptive filtering, sequenced adaptive filters, etc. However, the apparent lack of success of this type of methods (see, e.g., [15] and [125] for some of their major drawbacks) prompted a complete reformulation of the problem, whereby attention would be paid to the fundamental aspects behind the biological problem in hand.

Effectively, in [86] it is shown that the problem can be reformulated in a more efficient manner by looking at the bioelectrical phenomena ruling the heart activity and the propagation of heartbeat signals across the body (see also [15], [97], [102] and [134]). The result relies on an electrical model of the heart, the so-called *lead-vector* concept, which was introduced by Burger and Van Milaan as early as in the mid-forties [105], [106]. According to this model, at a certain distance from the heart its electrical activity can be represented by an equivalent current dipole with fixed position but variable magnitude and orientation. Respective dipoles are thus associated to the mother's and the fetus' heart. The geometrically orthogonal components of the dipoles, also called *heart vectors*, are in addition statistically independent, thus spanning three-dimensional signal spaces, the maternal and the fetal ECG signal subspaces.[1] These components constitute the *source signals* of the problem. The sources may be regarded as statistically independent owing to the physical independence of the underlying bioelectrical phenomena. The second major assumption is that over the frequency range of interest (viz., between about 0.5 Hz and 100 Hz) the body as a conductive medium behaves in a heterogeneous, resistive and linear manner. Further, since the heart signals can be considered as narrowband, propagation delays may be neglected. In conclusion, each of the p electrodes located on the patient's body, say $\mathbf{y}(k) = [y_1(k), y_2(k), \ldots, y_p(k)]^{\mathrm{T}} \in \mathbb{R}^p$, outputs an instantaneous linear combination of the bioelectric current sources. Assuming that the activity of all internal bioelectric current sources can be modelled by means of q unobservable independent *source signals*, $\mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_q(k)]^{\mathrm{T}} \in \mathbb{R}^q$, the electrode measurement vector accepts the matrix form:

$$\mathbf{y}(k) = M\mathbf{x}(k). \tag{6.1}$$

The structure of the mixing or transfer matrix $M \in \mathbb{R}^{p \times q}$ is determined by the body geometry, the electrode and source positions and the conductivity of the body tissues

[1] Strictly speaking, it has been evidenced [102] that the dimension of the FECG subspace is subject to variations during pregnancy, and can often be lower than 3.

[125]. Equation (6.1) corresponds to the familiar BSS model of instantaneous linear mixtures (Section 2.1). As an immediate consequence, BSS techniques may be applied to tackle the FECG extraction problem. The reader is referred to, e.g., [86] for a more exhaustive discussion on the theoretical justifications for the validity of model (6.1) in this particular biomedical application.

The methods described in [15] and [125] rely on the SOS of the data, seeking the removal of second-order dependencies in the observations (PCA, see Sections 2.1.5 and 2.2.1). Their major shortcoming is that the separation quality highly depends on a careful electrode placement selection. By contrast, exploiting the HOS and looking for higher-order independence at the separator output (ICA, see Sections 2.1.5 and 2.2.2) surmount this obstacle.

Various results obtained from the application of ICA-BSS methods to the biomedical problem in hand have already been reported in the literature [7, 33, 86, 97, 134, 136]. Also, they have been compared to other PCA-BSS methods [7, 86, 134, 136]. In this chapter we evaluate inside this biomedical problem some of the separation procedures developed in the previous chapters. In addition, in order to evidence that HOS-BSS based FECG extraction is indeed advantageous relative to other sort of techniques, we also compare these BSS methods with one of the most significant classical techniques proposed to solve this challenging problem: Widrow's multi-reference adaptive noise cancelling (MRANC) method [129].

6.3 Real ECG Recordings

Figure 6.1 displays the real cutaneous electrode recordings used in the experiments.[2] The signals were recorded over 10 seconds at a a sampling frequency of 500 Hz, yielding a total of $T = 5000$ samples per signal, from 8 cutaneous electrodes located on different points of a pregnant woman's body. Specifically, the first five recordings correspond to electrodes located on the mother's abdominal region. In them a mixture of FECG, MECG and noise is visible. The last three signals were digitized from the mother's thoracic region, and no FECG heartbeat component can be perceived, due to the

[2]The authors are grateful to L. De Lathauwer, D. Callaerts and J. Vandewalle, from K. U. Leuven, Belgium, for providing the ECG recordings.

longer distance between these electrodes and the fetal heart.[3] Regarding the vertical axes, only the relative amplitudes are important.

The choice of abdominal and thoracic (chest) electrode positions is justified by the fact that the maternal heartbeat constitutes the strongest interference. Then it seems reasonable to look for clear MECG components so that they can be subtracted, after suitable processing, from the abdominal leads, leaving only the desired fetal heart components. Additionally, the sought MECG signals must be as free as possible from FECG contribution, in order to prevent the latter from being cancelled out in the abdominal leads when performing the subtraction; hence the chest leads. This electrode placement strategy is critical for the MRANC method. By contrast, we anticipate (Section 6.6) that ICA-BSS methods are more robust to the electrode location, to the extent that thoracic leads are, as a matter of fact, dispensable.

6.4 Methods

Section 6.4.1 explores the extension of the two-signal BSS methods to scenarios of more than two signals. The basic aspects behind the MRANC method are then reviewed in Section 6.4.2.

6.4.1 Blind source separation: pairwise extensions

General pairwise procedure

Relying on the assumption that the surface potentials are generated according to model (6.1), the reconstruction of the FECG contributions to the recordings is reduced to the identification of the independent sources of fetal cardiac activity and the corresponding elements of the mixing matrix. This goal requires, firstly, the estimation of the source signals $\mathbf{x}$ and the mixing matrix M via a BSS method. In these experiments, we consider a Jacobi-like extension of the analytical methods which have originally been developed for the two-signal case in the previous chapters.[4]

[3] Further information on these recordings was not available.

[4] A priori, there is no reason why any of these particular methods should be preferred over other BSS procedures to deal with the FECG extraction. Our choice is merely based on illustrative and testing purposes, and motivated by the fact that other different ICA-BSS methods have already been applied to this problem, so that comparisons may be established from the new results presented in this

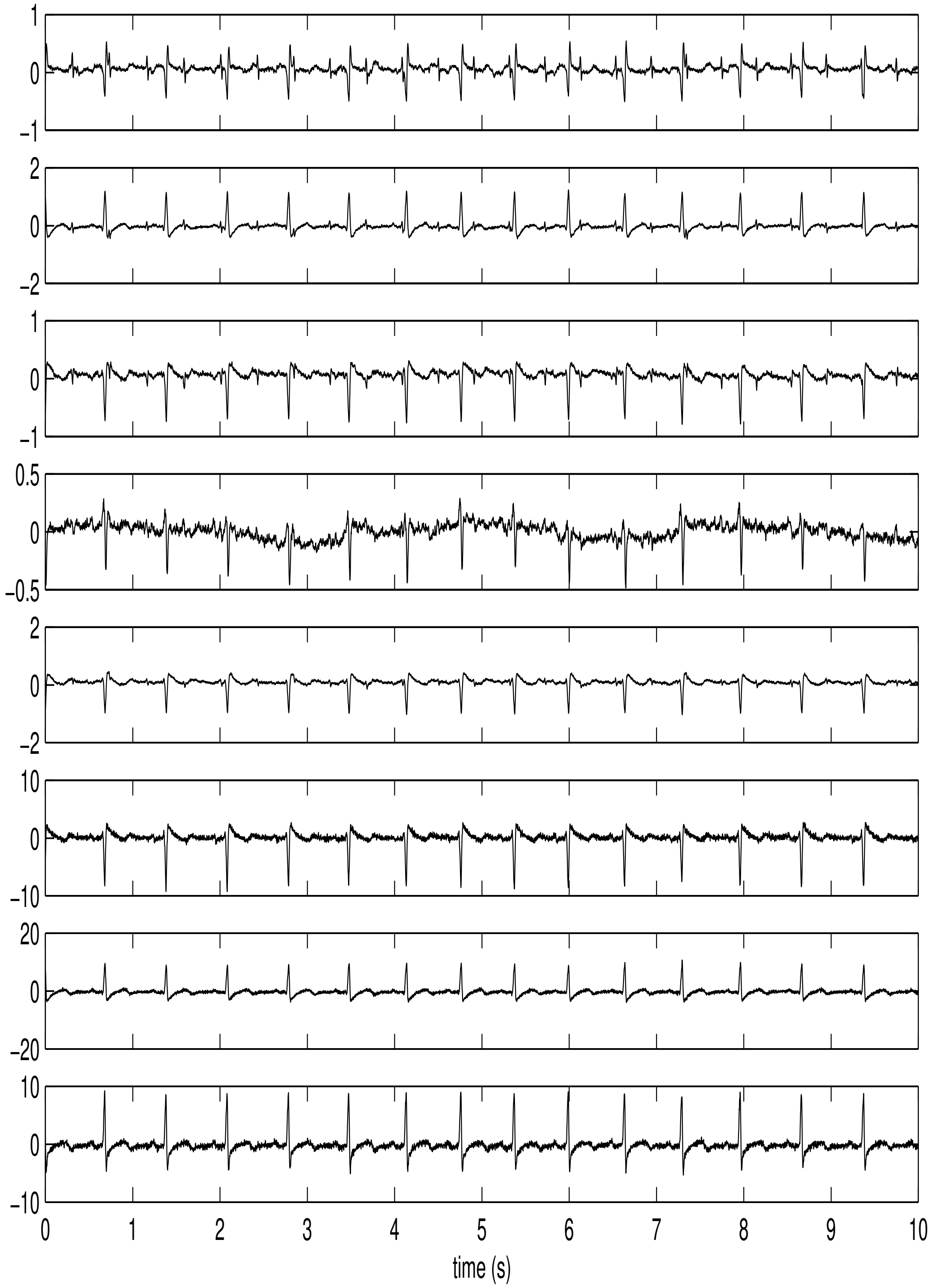


Figure 6.1: An 8-channel cutaneous electrode recording from a pregnant woman.

The extension, inspired by [45, 48] is made as follows. First, the observed inputs are processed in an initial PCA stage, via the SVD of the observation matrix. This is the pre-whitening step. The whitened signals are then processed in pairs, seeking higher-order independence. Elementary planar Givens transformations are set up by means of the respective closed-form estimators, and applied to each whitened signal pair in succession. Geometrically, this corresponds to performing at each iteration an elementary planar rotation on the two-dimensional scatter-plot of the current signal pair. A loop over all $p(p-1)/2$ signal pairs is called a *sweep*. Sweeps are repeated until convergence. This general pairwise-processing algorithm, which is summarized in Figure 6.2, was originally adopted by the HOEVD [48] to deal with more than two signals. JADE also operates in a similar pairwise fashion [23].

Sweeping strategies

The excursion across the signal pairs can be accomplished in various manners, as discussed in [45], comprising natural cyclic sweeping, dominant-element sweeping, and threshold sweeping. The first method takes the pairs in a pre-established order, and is the simplest. The second is relevant when contrast optimization is involved, and is reminiscent of the hermitian-matrix diagonalization procedure originally proposed by Jacobi [66]. The pair for which the objective function is the highest is given priority. In the third sweeping scheme, only the pairs for which the objective function is above a certain threshold are treated. Dominant-element sweeping is known to speed up convergence in many cases. However, the selection of each pair requires the calculation of the cost function for all pairs, so that computational cost is sacrificed [48]. For this reason, and for the sake of simplicity, we adhere to the natural cyclic sweeping strategy.

Preliminary experimental results

Several experiments on synthetic signals were performed as a preliminary evaluation of the EML, AEML and combEML extensions. To measure convergence, we defined the *absolute-rotation sum (ARS)* at sweep c as the sum over all signal pairs (i, j) of the

---

chapter.

Step 1. Obtain whitened signals from given observations (PCA). Consider the whitening-matrix (pseudo)inverse and the whitened signals as a first estimate of the mixing matrix and the sources, respectively: $\widehat{B} \rightarrow \widehat{M}$, $\mathbf{z} \rightarrow \hat{\mathbf{x}}$.

Step 2. For all signal pairs $\hat{\mathbf{x}}_{ij} = [\hat{x}_i,\, \hat{x}_j]^{\mathrm{T}}$:

2.1. Set up a (2×2)-unitary transformation $\widehat{Q}_{ij}$ from angle $\hat{\theta}_{ij}$ obtained via a suitable estimator (with $[z_1,\, z_2]^{\mathrm{T}} = [\hat{x}_i,\, \hat{x}_j]^{\mathrm{T}}$).

2.2. Counter-rotate: update $\hat{\mathbf{x}}_{ij} := \widehat{Q}_{ij}^{\mathrm{T}} \hat{\mathbf{x}}_{ij}$, $\widehat{M}_{:\,[i,\,j]} := \widehat{M}_{:\,[i,\,j]} \widehat{Q}_{ij}$, where $\widehat{M}_{:\,[i,\,j]}$ denotes the matrix composed of only the ith and jth columns of $\widehat{M}$.

Step 3. Repeat from Step 2 until convergence.

Figure 6.2: General algorithm for BSS of more than two signals via pairwise processing.

absolute values of the estimated rotation angles at that sweep, $\hat{\theta}_{ij}^{(c)}$:

$$\mathrm{ARS}(c) \triangleq \sum_{i<j} \big|\hat{\theta}_{ij}^{(c)}\big|. \tag{6.2}$$

It was observed that the ARS tended rapidly to zero as the number of sweeps increased, provided each source pair fulfilled the applicability conditions of the respective angle estimators. Numerous simulations indicated that convergence is usually reached within around $1 + \lfloor\sqrt{p}\rfloor$ sweeps,[5] which coincides with the value originally proposed for the HOEVD method [48]. Also, the number of AEML rotations carried out in each sweep by the combEML converged to the number of source-signal pairs with different kurtosis sign, just as expected from the practical decision rule derived in Chapter 5. On the part of the AEML (on its own), it was noticed that separation was generally satisfactory as well. However, convergence was hindered when some of the source signals had nearly equal kurtosis. In such situations, more sweeps were needed to reach convergence or, alternatively, given a fixed number of sweeps, worse separation quality was observed relative to cases of sources with different kurtosis values. Similarly, in occasions the decision rule of the combEML was observed to make the wrong estimator selection when one of the sources of the current pair had kurtosis near zero, because in that case the absolute value of the sks and the skd became very close.

[5]Recall that, in agreement with the comments made at the end of Section 2.2.1, the number of sources is assumed to be the same as the number of observations, $q = p$.

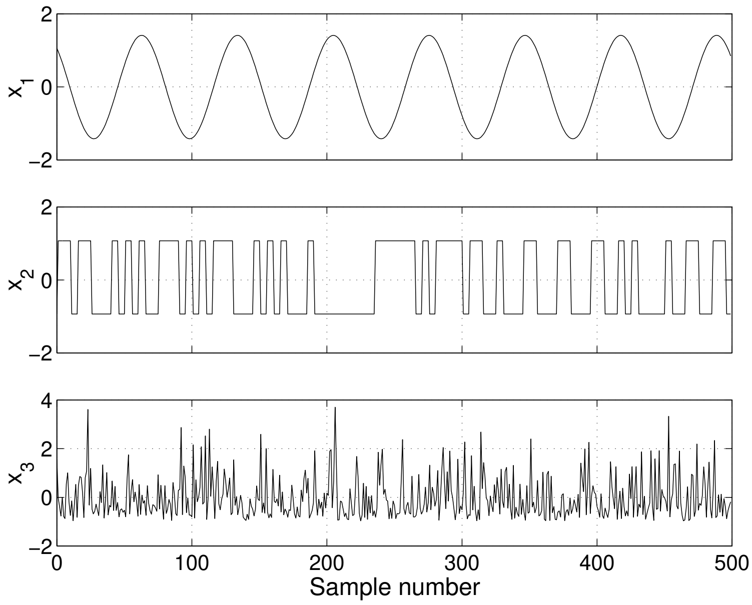


Figure 6.3: A realization of three sources: a sinusoid, a binary sequence and a white exponentially distributed random process.

Figures 6.3–6.6 serve as an illustration of the typical results obtained in these preliminary simulations on synthetic data. Fig. 6.3 displays a realization of three sources: a sinusoid, a binary sequence and an exponentially distributed signal, all composed of 1000 samples of which only the first 500 are shown. The three-sensor output resulting from a particular instantaneous linear mixture of those three sources appear in Fig. 6.4. The pairwise extension of the EML procedure supplies the sources shown in Fig. 6.5, whereas Fig. 6.6 displays the sources obtained by the HOEVD method.[6] Both sets of estimated source waveforms are nearly identical, apart from the different signal arrangements and unitary scale factors.

Computational complexity

The pairwise extensions of the EML, AEML and combEML procedures need $\mathrm{O}(p^{5/2}T)$ flops, for T large, to accomplish the source separation from a p-sensor mixture. This

[6]Original MATLAB code implementing HOEVD was downloaded from the URL: `http://www.i3s.unice.fr/~comon/matlab.html`.

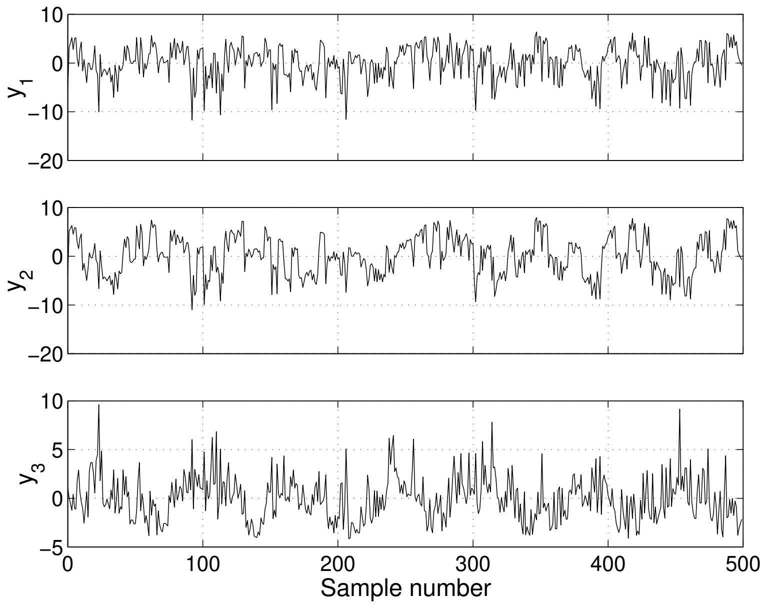


Figure 6.4: Instantaneous linear mixture of the source signals shown in Fig. 6.3.

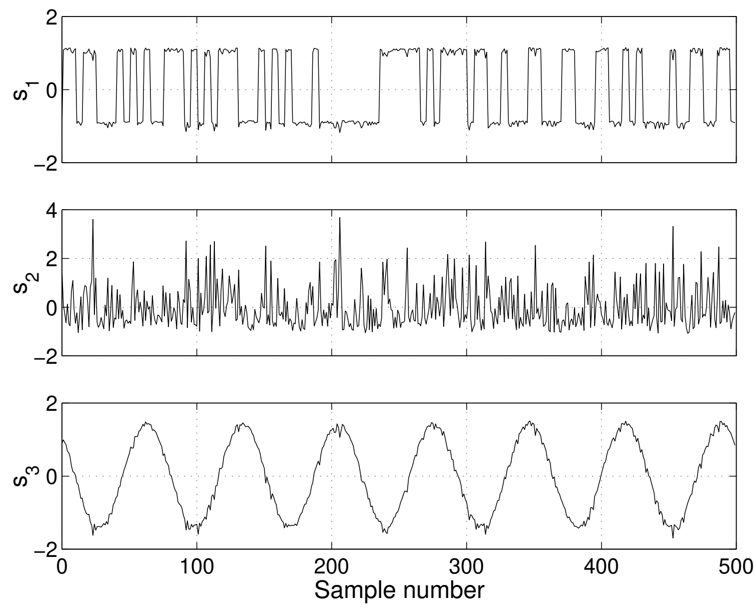


Figure 6.5: Sources retrieved by the EML method from the mixture shown in Fig. 6.4.

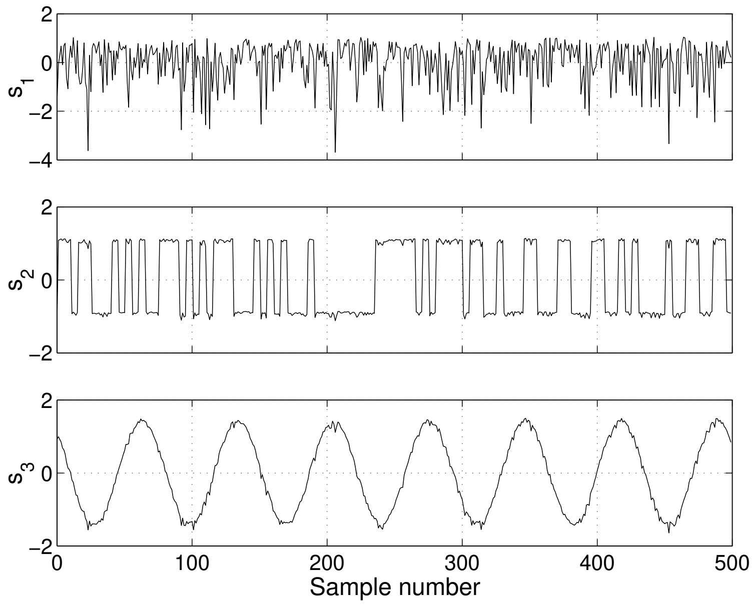


Figure 6.6: Sources extracted by the HOEVD method from the mixture shown in Fig. 6.4.

figure is easily determined from the cost analyses of Sections 3.3.6 and 5.2.5, and bearing in mind the typically required $1 + \lfloor\sqrt{p}\rfloor$ sweeps over the $p(p-1)/2$ signal pairs.

This value may be compared to the complexity of the two other BSS methods that will also be used in the experiments that follow, Comon's HOEVD [48] and Cardoso's JADE [23], which were described in Sections 2.5.4 and 2.5.5, respectively. First recall that both methods also process the whitened signals in a pairwise fashion, so that it is meaningful to compare them to the methods developed earlier in this section. HOEVD's cost is analyzed in [45, 48], and is the same as above. JADE operates directly on the cumulant matrices, so that the current signal-pair cumulants do not need to be recalculated from all signal samples at each iteration. This results in an extremely efficient implementation where the computational requirements at each sweep are independent of the sample size.

6.4.2 Multi-reference adaptive noise cancellation

One of the first successful approaches to the FECG extraction problem was developed by Widrow and colleagues in 1975. In their celebrated paper [129], an adaptive filtering standpoint was adopted. An abdominal lead, mainly containing a mixture of FECG and MECG signals, acts as a primary input to an adaptive noise canceller (Figure 6.7). The MECG interference that corrupts the abdominal leads is considered as the 'noise' to be eliminated. The reference inputs to the canceller are thoracic leads, mostly composed of maternal heartbeat components. The reference inputs are adaptively filtered and subtracted from the primary input, the adaptation criterion being to minimize the power of the output signal.

Mathematically, the MRANC method can be described as follows. Let $\{r^{(i)}(k),\, i = 1, \ldots, n\}$ be the n chest leads, acting as reference inputs (Fig. 6.7 shows the case $n = 3$). Let us define the associated vectors $\mathbf{r}_k^{(i)} \triangleq [r^{(i)}(k),\, r^{(i)}(k-1),\, \ldots,\, r^{(i)}(k-N+1)]^{\mathrm{T}}$, $i = 1, \ldots, n$, which contain N consecutive samples of such signals. These inputs are processed by means of the time-varying filters $\{w^{(i)}(k),\, i = 1, \ldots, n\}$, assumed to be finite impulse response (FIR) filters of tap-length N. Accordingly, the vectors $\mathbf{w}_k^{(i)} \triangleq [w_0^{(i)}(k),\, w_1^{(i)}(k),\, \ldots,\, w_N^{(i)}(k)]^{\mathrm{T}}$, $i = 1, \ldots, n$, define the filter coefficients at time instant k. The primary abdominal lead is represented by $a(k)$. The output at instant k is then given by $\varepsilon(k) = a(k) - s(k)$, with

$$s(k) = \sum_{i=1}^{n} w^{(i)}(k) * r^{(i)}(k) = \sum_{i=1}^{n} \mathbf{w}_k^{(i)\mathrm{T}} \mathbf{r}_k^{(i)} = \sum_{i=1}^{n} \mathbf{r}_k^{(i)\mathrm{T}} \mathbf{w}_k^{(i)}, \tag{6.3}$$

where symbol $*$ stands for the convolution operator. Aiming to reduce the MECG disturbance at the canceller output, the filter coefficients are updated so as to minimize the output power, $\mathrm{E}[\varepsilon^2]$. However, this is impractical in a real time implementation. A more practical solution is attained via the steepest descent method over the instantaneous stochastic approximation of the true gradient. This yields the well-known least mean square (LMS) update [64,70,129,130], which in our multi-reference set-up reads:

$$\mathbf{w}_{k+1}^{(i)} = \mathbf{w}_k^{(i)} + 2\mu_k^{(i)} \varepsilon(k) \mathbf{r}_k^{(i)}, \qquad i = 1, \ldots, n. \tag{6.4}$$

Parameter $\mu_k^{(i)}$ is the adaption coefficient or step-size of the ith filter at time k. It poses a trade-off among convergence speed, accuracy and stability of the algorithm [129,130]. To improve its robustness with respect to the input power and speed up convergence, $\mu_k^{(i)}/\|\mathbf{r}_k^{(i)}\|^2$ can be substituted for $\mu_k^{(i)}$ in (6.4), giving rise to the normalized LMS (NLMS) algorithm [64, 70]. In a number of experiments $\|\mathbf{r}_k^{(i)}\|^2$ proves too crude an

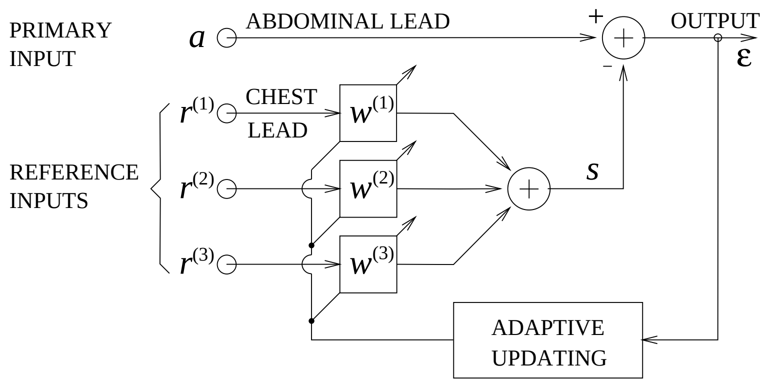


Figure 6.7: MRANC set-up for the FECG extraction problem.

approximation for $NP^{(i)}$, $P^{(i)}$ being the power of the ith reference signal. Instead, the adaptive estimate

$$\hat{P}_{k+1}^{(i)} = (1 - \tilde{\mu}_k^{(i)})\hat{P}_k^{(i)} + \tilde{\mu}_k^{(i)} r^{(i)}(k)^2, \tag{6.5}$$

provides smaller misadjustment than the NLMS as at a similar convergence rate and lower computational cost. This variant is a particular version of the so-called power-normalized LMS (PNLMS) algorithm [64].

Finally, straightforward calculations lead to a number of flops of $\mathrm{O}(nNT)$ for the MRANC to extract the FECG components from a single primary lead.

6.5 Experimental Methodology and Results

For Widrow's MRANC solution, the three thoracic leads of Fig. 6.1 are employed as reference inputs to the canceller (i.e., $n = 3$, as depicted in Fig. 6.7), whereas the abdominal leads play (in turn) the role of primary inputs. We first need to work out the number of filter taps, which, for simplicity and due to the identical nature of the reference signals, is chosen to be equal for all filters. Standard guidelines given in [130] yield $N = 60$ taps. This provides a total delay, with uniformly spaced taps, of 120 ms.[7] The other parameters needed by the PNLMS algorithm described in Section 6.4.2 are

[7] cf. 129 ms of the experiments reported in [129, Section VIII-C], where log-periodic tap spacing is employed.

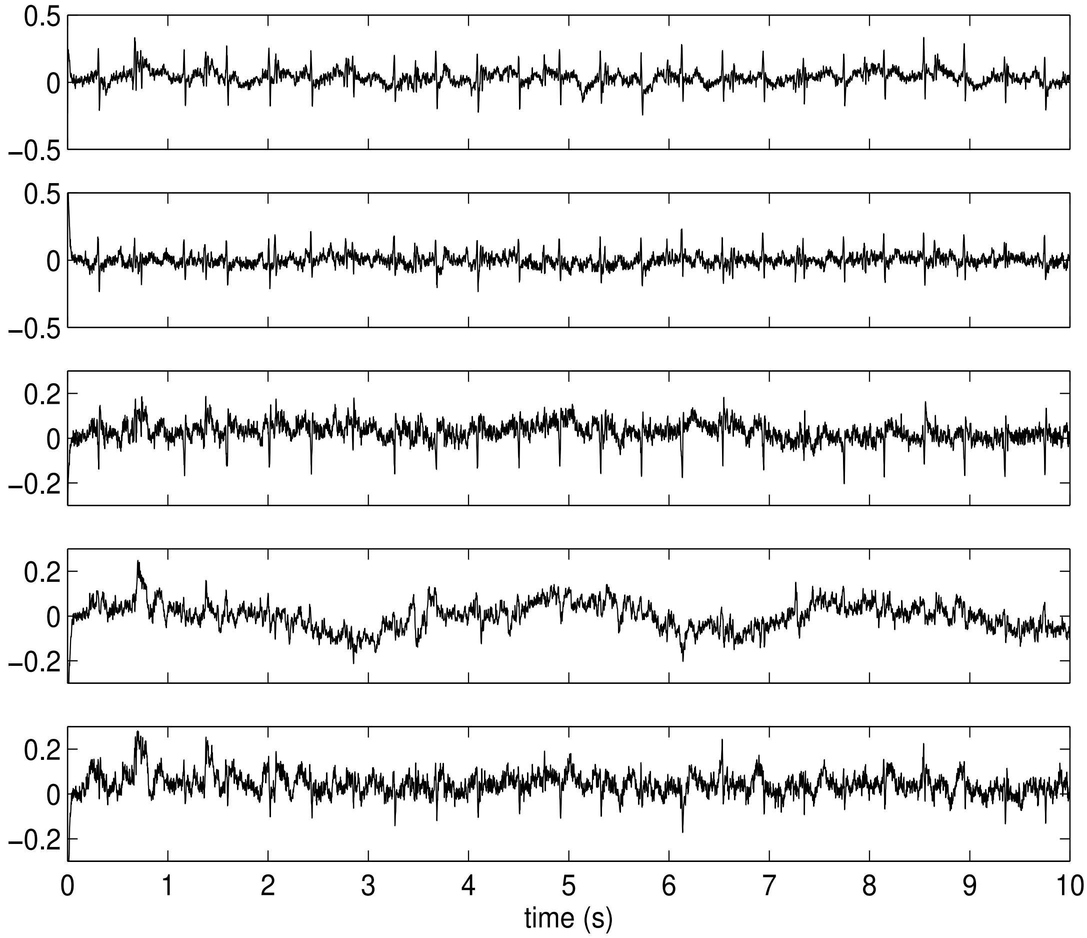


Figure 6.8: Fetal contributions to the abdominal electrodes (first five signals) of Fig. 6.1 obtained by the MRANC method.

selected as: $\mu_k^{(i)} = \mu = 10^{-2}$ and $\tilde{\mu}_k^{(i)} = \tilde{\mu} = 2 \cdot 10^{-3}$, $\forall i$, $\forall k$.[8] The resulting waveforms obtained from each of the five abdominal leads are shown in Fig. 6.8.

As to the BSS methods, the recordings of Fig. 6.1 were fed into the methods, in order for them to perform the source extraction and mixing matrix identification of the assumed BSS problem. Since both HOEVD and JADE[9] make direct use of the 4th-order cumulants, the closed-form angle estimation in the EML, AEML and

[8] The choice of $\tilde{\mu}$ ensures convergence of (6.5) to the true power value around $k = T/2$ for stationary signals.

[9] Original MATLAB code implementing JADE was downloaded from the URL: `http://www-sig.enst.fr/~cardoso/guidesepsou.html`.

combEML was carried out through the whitened-signal cumulant forms, rather than the equivalent scatter-plot based forms. As an indication of the convergence characteristics, Figure 6.9 shows the ARS [eqn. (6.2)] by each method after consecutive sweeps. In total, 5 sweeps were performed by all methods. The final absolute rotations (after the 5th and final sweep) are displayed in Table 6.1. That table also shows the degree of 4th-order independence of the recovered sources (index $\tilde{\Upsilon}_4$, see Section 2.4.4).

The waveform similarity (Section 2.4.1) between the source signals estimated by the different methods, included the output of the PCA step only, is summarized in Table 6.2. The sources estimated by all the higher-order methods were visually very similar, and only the AEML presented certain significant differences. As an illustration, Figure 6.10 shows the sources retrieved by the combEML.

For the combEML, the number of AEML rotations reached a stationary value of 16 from the 4th sweep. The normalized kurtosis values of the sources estimated by the combEML were: 28.61, 27.54, 0.16, 8.08, 8.24, -0.64 , 7.18, and 0.16. Hence, in compliance with rule (5.40), only 7 AEML rotations were expected. It was found out that in pairs (1, 3), (1, 8), (2, 3), (2, 4), (2, 8), (3, 7), (4, 8), (5, 8) and (7, 8), the AEML was wrongly selected (i.e., the sks was actually bigger than the skd, in absolute value, yet the opposite was being estimated). In all these pairs, apart from (2, 4), a source of near-zero kurtosis appeared, which accounts for the respective errors (as commented in the preliminary results reported at the end of Section 6.4.1). The decision error in pair (2, 4) was due to the fact that those source signals were found to be relatively high correlated at 4th-order (which, in turn, was connected to the fact that both signals belonged to the MECG subspace; see below). We finally realized that using $\hat{\gamma}$ instead of $\hat{\xi}_{\mathrm{EML}}$ to work out the absolute value of the sks, the results became as expected: the type of all signal pairs was identified correctly, and 7 AEML rotations were being performed right from the 2nd sweep, in agreement with the decision rule and the values of source kurtosis. This subtle variation, hence, proves more robust in quasi-Gaussian (i.e., with near-zero kurtosis) and partially correlated source environments, as substantiated through a number of further simulations on synthetic data.

With this new version of the combEML method, the angle trajectory changed slightly, as depicted in Figure 6.11. Not only the expected number of AEML rotations were now performed, but also the absolute AEML rotations per sweep decreased dramatically relative to the original decision rule. However, the total angle trajectory remained basically unaltered, and so did the convergence. The final absolute rotation was now 0.36°.

Since most rotations were now executed with the EML estimator (Fig. 6.11), the new resulting sources by this combEML variant became almost identical (even in ordering and scale) to those by the EML (waveform similarity: 99.98%). The EML sources are represented in Figure 6.12. Still, these are very similar to the original combEML sources of Fig. 6.10; precisely, only the 2nd and 4th waveforms show discernible differences.

In all cases, 3 MECG and 2 FECG source signals [respectively, waveforms (1, 2, 4), and (5, 7) in Figs. 6.10 and 6.12; this source ordering was different for the AEML and JADE] were identified among the set of estimated sources. The latter set, which span the FECG subspace, were isolated in vectors $\hat{\mathbf{x}}_f = [\hat{x}_{f_1}, \hat{x}_{f_2}]^{\mathrm{T}}$. Matrix $\widehat{M}_f = [\hat{\mathbf{m}}_{f_1}, \hat{\mathbf{m}}_{f_2}]$ contained the corresponding columns of the estimated mixing matrix. The fetal heartbeat contributions to the recordings were then calculated through:

$$\hat{\mathbf{y}}_f = \widehat{M}_f \hat{\mathbf{x}}_f. \tag{6.6}$$

The FECG contributions to the abdominal recordings obtained by the combEML and the JADE appear in Figures 6.13 and 6.14, respectively. Visually, they are identical. Indeed, the similarity between the waveforms was 99.6%. Observe that this processing on the fetal sources allows the estimation of FECG contributions to *all* leads, even the thoracic ones. Nevertheless, only the abdominal signals are shown here for the sake of a more meaningful comparison with the MRANC.

Further experimental results on FECG extraction via BSS are reported in a number of references. In [33], experiments with JADE on the dataset of [93] can be found; in [6], [7] and [134] the methods evaluated are the PCA, the HOEVD [48] and the HOSVD [85,86] on the dataset used in these experiments (the source signals estimated by those methods are displayed therein); in [140] both datasets are employed to test the EML, which is also compared to Widrow's MRANC.

The results presented in this section are discussed next.

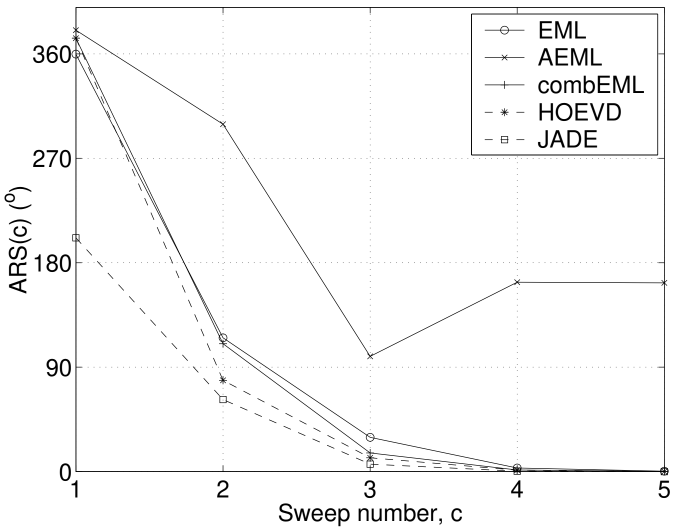


Figure 6.9: Absolute-rotation sum after successive sweeps in the processing of the recordings of Fig. 6.1.

EML	AEML	combEML	HOEVD	JADE
Absolute-rotation sum, ARS (°)				
0.42	162.62	0.21	0.11	0.03
Statistical independence at 4th-order, $\tilde{\Upsilon}_4$ (dB)				
−7.86	−5.90	−7.04	−8.25	−8.20

Table 6.1: Some performance measures after the 5th sweep in the processing of the recordings of Fig. 6.1.

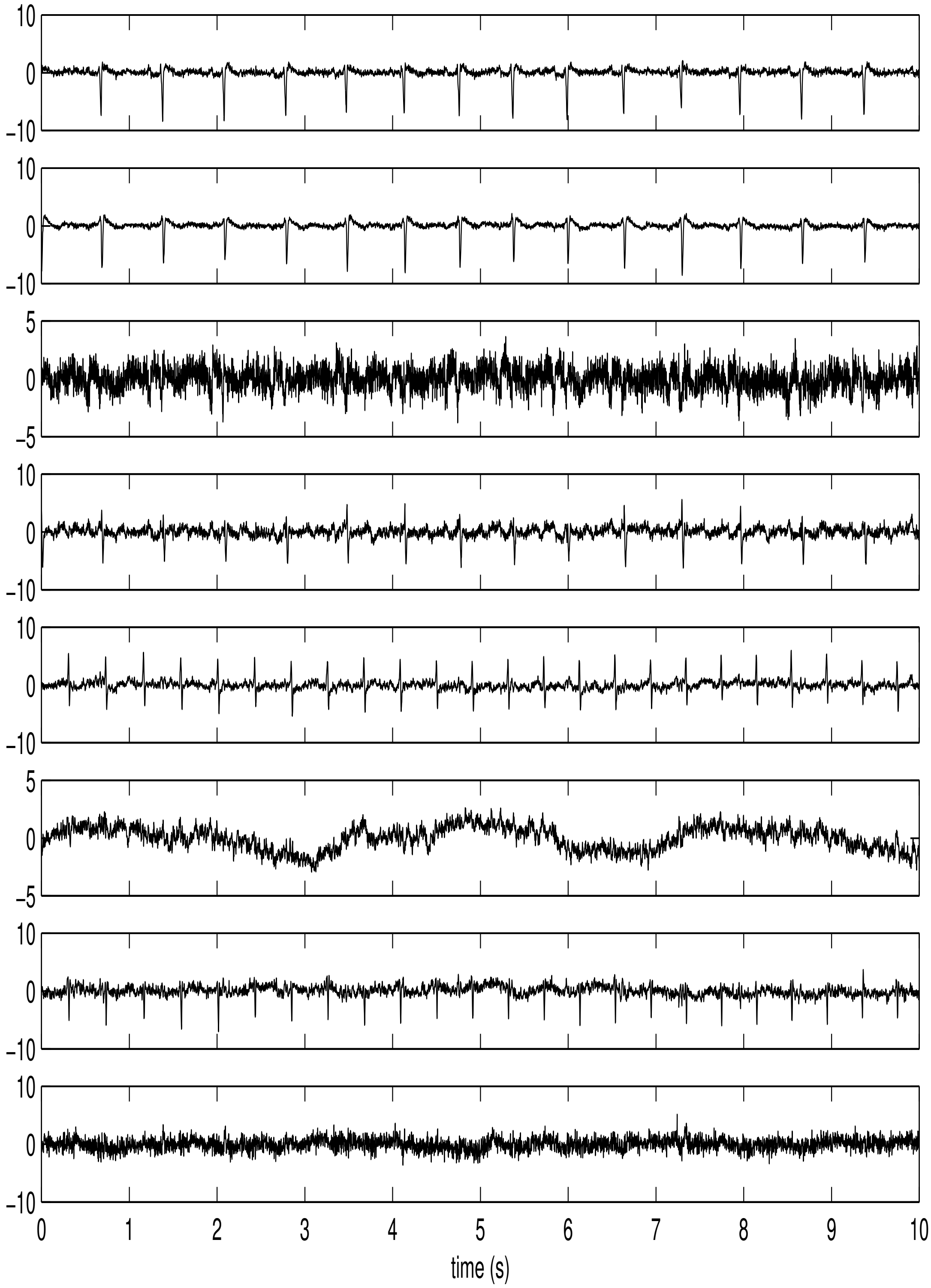


Figure 6.10: Source signals extracted by the combEML method from the ECG recordings of Fig. 6.1.

	PCA	EML	AEML	combEML	HOEVD	JADE
PCA	—	71.0	71.8	72.5	72.2	71.8
EML	71.0	—	80.3	97.6	99.7	94.3
AEML	71.8	80.3	—	84.8	81.3	87.8
combEML	72.5	97.6	84.8	—	98.8	93.5
HOEVD	72.2	99.7	81.3	98.8	—	94.2
JADE	71.8	94.3	87.8	93.5	94.2	—

Table 6.2: Waveform similarity, ε^2 (%), between the sources recovered by several methods from the recordings of Fig. 6.1.

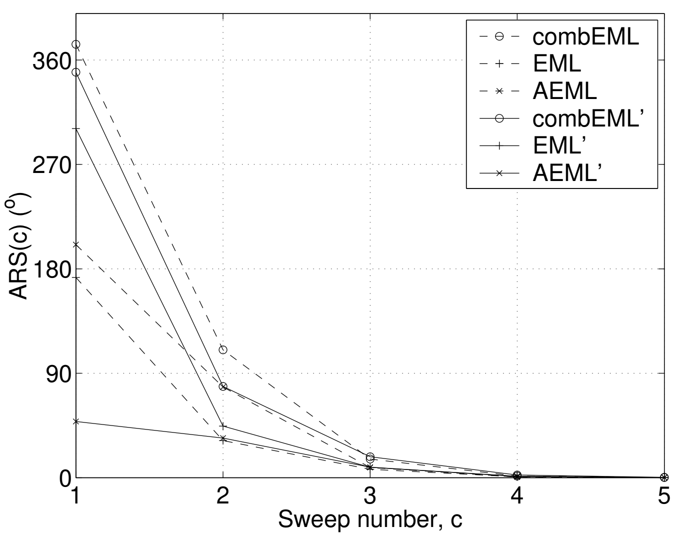


Figure 6.11: Total, EML and AEML absolute-rotation sums in combEML sweeps, in the processing of the recordings of Fig. 6.1. Dashed lines: decision rule (5.42) from EML centroid. Solid lines: modified decision rule, where a direct sks-estimate [based on eqn.(3.35)] is used instead.

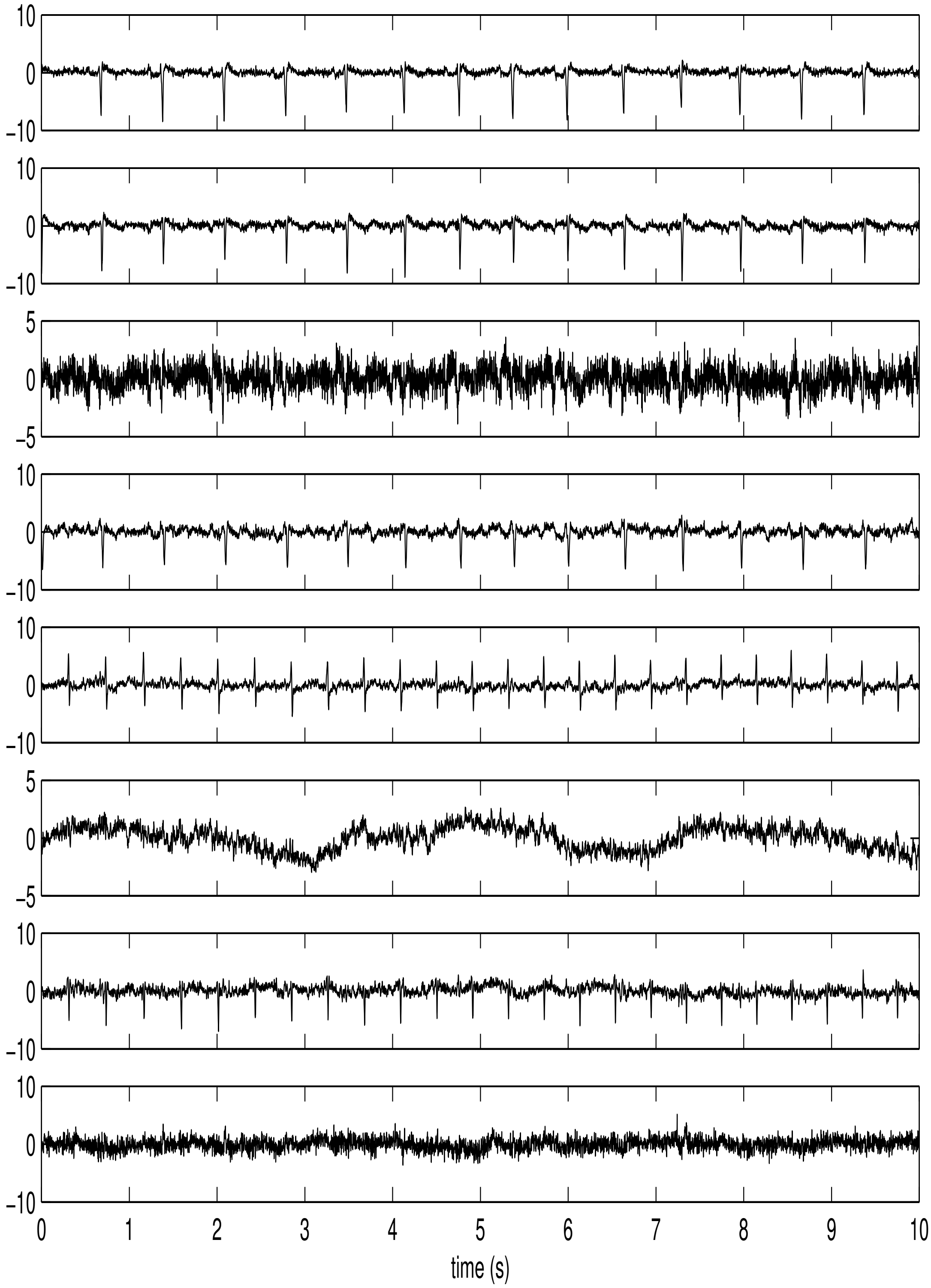


Figure 6.12: Source signals recovered by the EML method from the ECG recordings of Fig. 6.1.

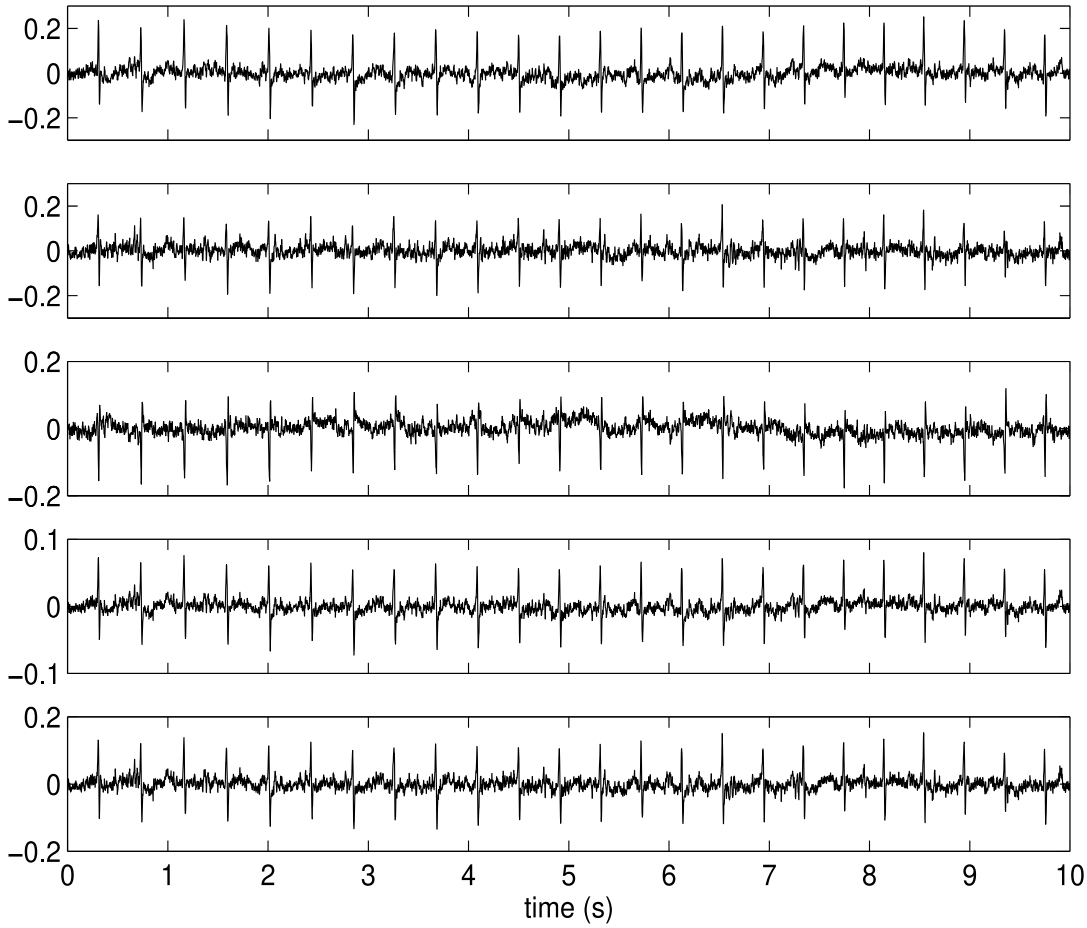


Figure 6.13: FECG contributions to the abdominal electrodes (first five signals) of Fig. 6.1 obtained by the combEML method.

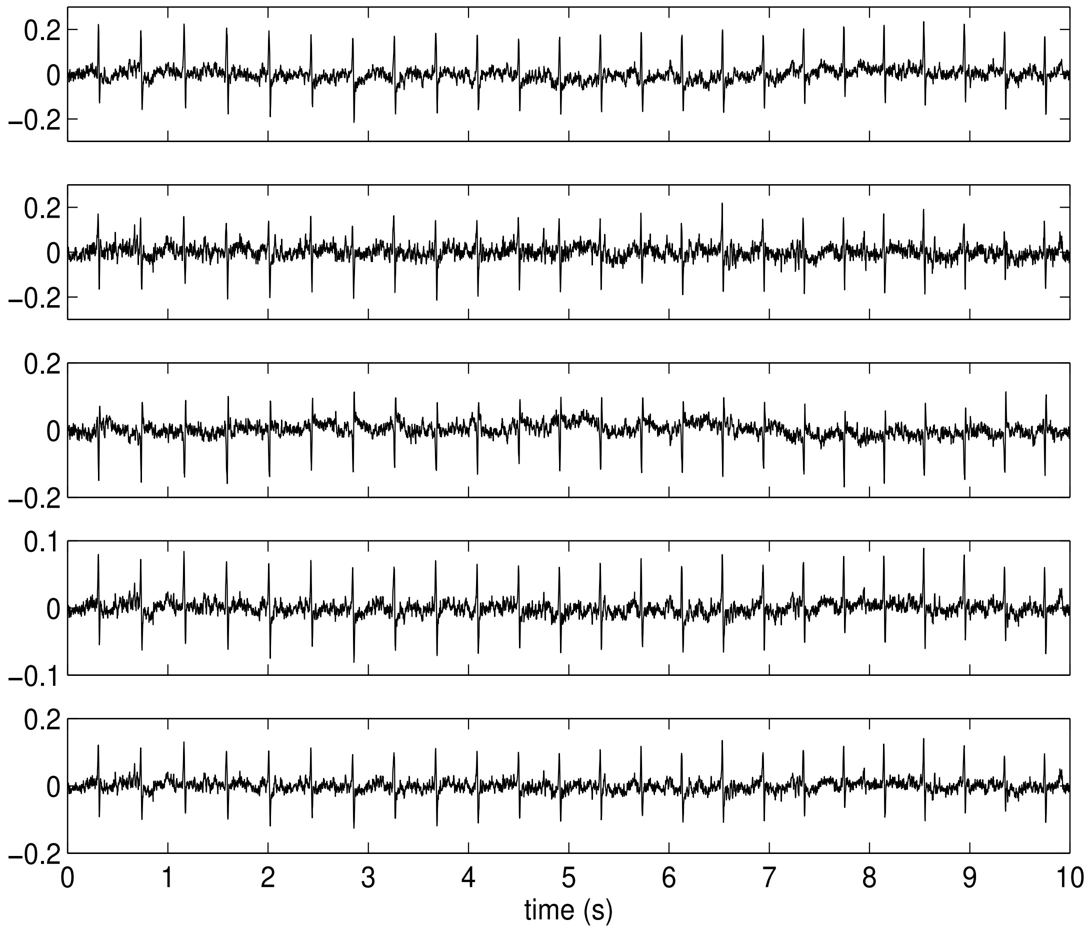


Figure 6.14: FECG contributions to the abdominal leads (first five signals) of Fig. 6.1 obtained by the JADE method.

6.6 Discussion

Out of the waveforms recovered by the MRANC method in Fig. 6.8, only the 1st and 3rd signals display fairly clear fetal heart components. MECG interference still corrupts the 2nd output, while the 4th signal is heavily corrupted by noise and a periodic component, presumably due to the mother's respiration. Since there is hardly any of this periodic component present in the chest leads, it cannot be filtered and subtracted from the 4th abdominal electrode, where it appears. The results from this approach, therefore, seem to be very dependent on the primary electrode placement, similarly as occurs with SOS-based BSS techniques [15, 125]. Totally analogous results are obtained for the dataset of [93], as reported in [140].

The quality of all the FECG contributions reconstructed by the BSS methods is notably superior. For instance, the periodical component which corrupted the 4th signal obtained by the MRANC solution (Fig. 6.8) is now completely eliminated in Figs. 6.13 and 6.14, since it is found to be generated by an independent source of interference and hence it is extracted from the 4th recording as a separate noise signal. Similar outcomes are obtained by the other BSS methods. The success in the extraction of the FECG contributions to the electrodes is directly related to the quality of the separation of the different source signal subspaces (maternal, fetal and noise), as can be appreciated in Figs. 6.10 and 6.12.

As cited earlier, the results of all five BSS methods considered are extremely similar, and indeed visually identical. Only the AEML shows certain discrepancies. In the results offered by this method, the MECG subspace was clearly revealed. The 1st and 2nd maternal sources were less noisy than the other methods, but the 3rd one was slightly more noisy. Similarly, the 1st fetal source was very close to that extracted by the other methods, but the 2nd one manifested greater noise interference and MECG contribution. The comparatively worse performance of the AEML is owing to some of the source signals having near-zero kurtosis (quasi-Gaussianity), where the AEML estimator shows difficulties (Section 5.2), thus impairing the convergence of its pairwise extension. This is endorsed by the cumulative angles shown in Fig. 6.9 and Table 6.1. By contrast, the other methods show a rapid stable convergence, JADE being the fastest, followed by the HOEVD, the combEML and the EML, respectively.

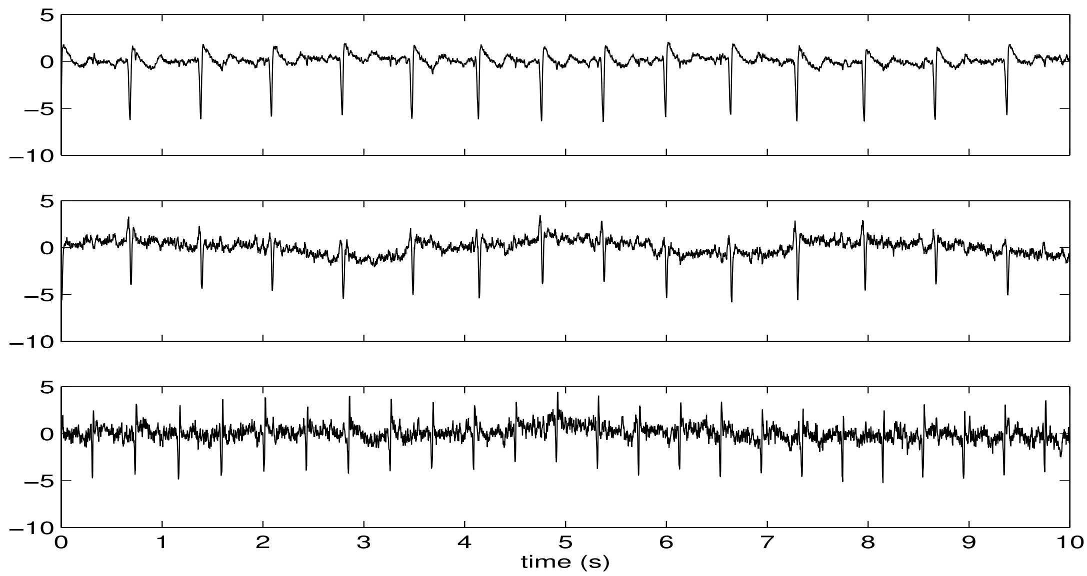


Figure 6.15: Source signals recovered by the combEML method from only 3 abdominal electrodes (waveforms 3, 4 and 5) of the ECG recordings of Fig. 6.1. Three sweeps over the signal pairs were performed.

The BSS robustness with respect to the electrode placement was evidenced through further experiments whereby FECG sources were clearly revealed by exploiting abdominal leads exclusively. Favourable results were obtained by processing the output of up to only 3 electrodes, as illustrated in Figs. 6.15 and 6.16. In Fig. 6.15, the fetal and maternal subspaces are successfully separated, and an FECG source (3rd signal) practically free from MECG interference is recovered. From the fetal source, the FECG contributions to the electrodes involved are obtained as explained in Section 6.5, and are displayed in Fig. 6.16 (cf. last 3 plots of Figs. 6.8, 6.13 and 6.14). The dimension of the estimated fetal subspace is now lower than in the full-recording processing, which accounts for the dissimilar results in both cases. However, the FECG extraction achieved from such smaller number of electrodes can still be considered as satisfactory.

The computational burden is another relevant issue. According to the figures presented in Sections 6.4.1 and 6.4.2, the relative cost between the MRANC and the BSS techniques, in general and under the same conditions and n of the order of p, depends on the number of taps in the MRANC filters. For the parameter values of these particular experiments, however, the BSS methods were more expensive.

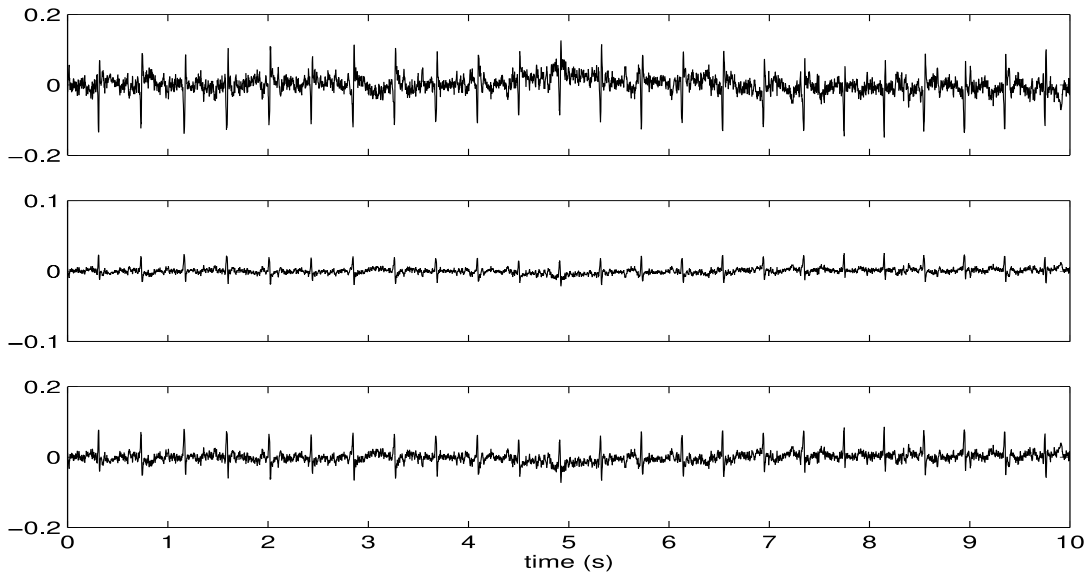


Figure 6.16: FECG contributions to electrodes 3, 4 and 5 reconstructed from the source signals of Fig. 6.15.

It must also be remarked that the MRANC is adaptive (on-line) by nature, whereas the BSS method processes the data in sample blocks (off-line or batch processing). For the application in a clinical environment, on-line processing is certainly more convenient. Adaptive BSS methods do exist as well (see, e.g., [29, 56], or also the adaptive versions studied in Chapter 7), but additional experiments on these ECG data confirm that more samples are needed by these adaptive algorithms in order to reach a satisfactory stable solution. This is connected to the larger sample sizes required to compute HOS with a reasonable estimation accuracy. In the MRANC, by contrast, only SOS are (implicitly) used.

The relevance of the performance measures deserves some final comments. The source independence figures of Table 6.1 show that the HOEVD offers the best performance (the lowest $\tilde{\Upsilon}_4$). However, from the visual inspection of the recovered sources (comparing the HOEVD sources in [7,134] and Figs. 6.10 and 6.12) and from the source waveform similarity values of Table 6.1, this outcome is not apparent. Since the HOEVD is specifically designed to maximize the degree of 4th-order independence precisely as it is measured by the parameter $\tilde{\Upsilon}_4$, this method presents an 'unfair' advantage when assessing the performance through such index. A lower $\tilde{\Upsilon}_4$, hence, does not necessarily imply a superior separation quality, at least in this particular application. This exemplifies the remarks of Section 2.4.6.

6.7 Summary and Conclusions

The two-signal methods studied in the previous chapters of this thesis have been extended to the general BSS scenario composed of more than two sources and two measurements. The extension is based in the pairwise processing notion first proposed in [45,48], and entails a computational cost of the same order as the method described therein (the HOEVD). The whitened signals are processed in pairs via any of the two-signal closed-form estimators, and sweeps over all pairs are repeated until convergence. Although the convergence of this procedure is still mathematically unproven, it has been reached in all performed experiments, as long as the sources fulfilled the validity conditions of the particular analytic estimator(s) being used.

As an illustration of the validity of this extension, some of the methods have been applied to the relevant biomedical problem of the fetal ECG extraction, providing excellent performance. The fetal heartbeat contributions to a real corrupted ECG recording taken from a pregnant woman have been neatly separated from other sources of disturbance. Results from the closed-form methods have been virtually identical to those obtained by contrast-based techniques such as HOEVD and JADE. As a by-product, the experiments have allowed us to realize a subtle but important variation on the practical decision rule derived in Chapter 5. By this modification, the rule is made more robust to separation environments with quasi-Gaussian and partially correlated sources.

Also, a comparison has been made with a classical method for FECG extraction, Widrow's MRANC. The conclusion is that HOS-based BSS proves a more robust, successful approach to the non-invasive FECG extraction problem than the MRANC, although the superior performance is attained at the expense of an increased computational complexity. Nonetheless, the achieved FECG-extraction quality offers promising prospects for the use of ICA-BSS techniques in prenatal medical diagnosis.

Chapter 7

Adaptive Separation

7.1 Introduction

In certain applications it is desirable, or indeed required, to accomplish the source signal extraction and mixing matrix identification in real time, which calls for adaptive (or on-line) BSS implementations. By adaptive is understood a processing whereby the estimation of the source signals and the mixing matrix is carried out in a sample-by-sample fashion, as the observation samples are being produced at the sensor output. This contrasts with batch processing, where a whole block of input samples is processed at a time. For instance, in the biomedical problem of fetal ECG extraction, introduced in the previous chapter, such adaptive methods would allow on-line monitoring and condition diagnosis of the fetus' heart. In other situations, the parameters of the mixing system or the source signals themselves are not stationary, but are subject to variations with time. A typical example is a multi-user mobile communication system [29], where those variations in the mixing pattern occur as the users roam around, thus changing the channel characteristics. Even if the users remain still, the channel is still subject to time variations (fading) due to changing atmospheric conditions and environment. In all these situations, the performance of the so-called batch-processing or off-line methods would be very poor, and adaptive procedures are necessary for a reasonable separation performance.

Several adaptive methods for BSS can be found in the literature. Interestingly enough, the first adaptive implementation was also the first successful attempt to tackle the BSS problem [71, 75]. The adaptive version of the CF method is proposed in [41]. The CF estimator depends exclusively on the 4th-order cross-cumulants of the whitened sensor outputs [via ρ, see eqn. (3.74)]. Hence, its adaptive implementation reduces to a simple adaptive cumulant estimation, yielding the so-called *adaptive rotation (AROT)* algorithm. The adaptive method based on the contrast function independently developed in [48] and [62] is implemented in [68]. The major drawback of the implementations of [18] and [68] is that, by construction of their respective algorithms, it is difficult to carry out their convergence study.

Delfosse and Loubaton [56], undertake a more successful approach. Their adaptive technique is based on the contrast functions of Comon [48] and Moreau [95], plus two particular devices: a special parameterization of the unitary transformation and a deflation strategy, whereby the sources are extracted one by one, one after another. The deflation scheme presents the disadvantage that errors propagate and accumulate in the successive source extraction stages. However, the employed devices allow the global convergence of the algorithm to be proven analytically for i.i.d. source sequences.

A different approach is adopted in [29] by Cardoso and Laheld. Taking advantage of the implicit multiplicative structure of the BSS problem and employing the "relative gradient", they derive a family of adaptive algorithms with serial updating, i.e., the separating matrix is updated at each iteration by multiplying the one obtained at the previous iteration by another matrix. This class of algorithms — so-called *equivariant adaptive separation via independence (EASI)* — is shown to exhibit uniform performance characteristics [25, 27, 29, 84], that is, the quality of the source estimation does not depend on the particular value of the mixing matrix. In contrast to the methods of [41], [68] and [75], the EASI performance analysis is easily accomplished with the conventional tools for the study of stochastic algorithms. On the negative side, the stability of the algorithm is very dependent on the choice and tuning of certain non-linear functions with respect to the source statistics, which in a genuine blind separation problem are not known a priori. Also, its convergence to non-spurious solutions is not guaranteed.

We derive in the present chapter adaptive algorithms that avoid those two drawbacks, namely, the dependence of their convergence on the source statistics (shown, e.g., by the methods of [29] and [75]), and the difficulties in analyzing their asymptotic performance (as seen, e.g., in the methods of [41] and [68]). By contrast, the procedures considered here exhibit two very desirable properties: 1) convergence (virtually) independent of the source distribution, and 2) straightforward performance analysis. The algorithms are based on the batch-processing method studied in the previous chapters, from which the first property is inherited. From these 4th-order off-line procedures, their adaptive counterparts are derived, resulting in simple updating rules in the typical form of stochastic adaptive algorithms, so widely studied [11]. The analysis of their convergence and asymptotic properties is easily carried out, and is presented herein as well. In principle, most of the derivations are developed in the simplified noiseless two-source two-sensor BSS scenario, but it is also shown how the procedure can be extended to more than two signals, in the pairwise processing fashion already suggested in the literature (for instance, in [41] and [48]) and used for batch processing in Chapter 6. Simulations endorse the effectiveness of this extension.

The chapter is organized as follows. In the first place, Section 7.2 describes mathematically the BSS problem in its adaptive framework. Next, Section 7.3 develops the adaptive version of the EML method, its extension to more than two signals being implemented in Section 7.5. The asymptotic performance and convergence characteristics of the algorithm in the two-signal case are analyzed in depth in Section 7.4. The adaptive extension of the combined estimation scheme developed in Chapter 5 is accomplished in Section 7.6. Section 7.7 studies the computational complexity of the algorithms. Section 7.8 reports on some simulation experiments run in order to test the validity of the theoretical results, as well as to compare the proposed methods with other adaptive implementations. Concluding remarks in Section 7.9 bring this chapter to an end.

7.2 Adaptive Blind Separation

The potential channel variability in non-stationary environments is now introduced in the BSS model by means of a (possibly) time-varying mixing matrix M_k, $k = 1, 2, \ldots,$[1]

[1]For simplicity in the notation, in the sequel we will represent the time variable as a subindex when convenient; that is, we may write $\mathbf{x}_k$ for $\mathbf{x}(k)$, and so forth. Both expressions will be used indistinctly.

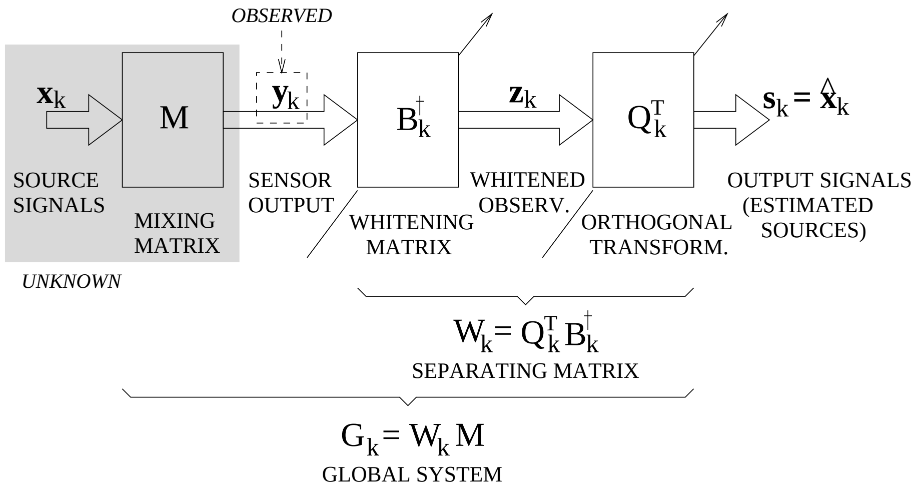


Figure 7.1: Schematic depiction of the adaptive BSS.

so that the noiseless model reads:

$$\mathbf{y}_k = M_k \mathbf{x}_k. \tag{7.1}$$

As for bath processing, the separation process can be accomplished adaptively in two steps. At each iteration k the whitening matrix and the unitary transformation, $\widehat{B}_k^\dagger$ and $\widehat{Q}_k$, are updated.[2] The first matrix aims to yield the whitened components:

$$\mathbf{z}_k = \widehat{B}_k^\dagger \mathbf{y}_k, \tag{7.2}$$

whereas the second is intended to provide an estimate of the source vector:

$$\hat{\mathbf{x}}_k = \mathbf{s}_k = \widehat{Q}_k^T \mathbf{z}_k. \tag{7.3}$$

The separation matrix is then given by the product $W_k = \widehat{Q}_k \widehat{B}_k^\dagger$, and the global matrix by $G_k = W_k M$. Figure 7.1 summarizes graphically the typical set-up for adaptive BSS, as well as the terminology and all the matrix and vector relationships referred to above, which are totally analogous to those employed in the batch-processing case.

The on-line estimation of $\widehat{B}^\dagger$ is called *adaptive pre-whitening*, and is accomplished through the adaptive counterparts of the SOS-based procedures described in Section 2.2.1.

[2] We make the iteration number coincide with the sample number, so that the processing takes place in a sample-by-sample fashion.

7.2.1 Adaptive pre-whitening

Several different methods exist for adaptive pre-whitening [29, 41, 56, 95]. In [41], a Cholesky factor [66] $\widehat{B}$ of the sensor-output covariance matrix R_y is adaptively estimated by means of a $(p+1) \times (p+1)$ matrix built up as $E_k = \begin{bmatrix} (1-\lambda_k)\widehat{B}_{k-1} & \lambda_k \mathbf{y}_k \\ 0 & 1/\lambda_k \end{bmatrix}$, λ_k being a sequence of positive adaption steps. These coefficients balance a trade-off between convergence speed and accuracy, as it is well established from the standard theory of adaptive systems [70, 130]. A sequence of p Givens rotations is found that cancel out, by post-multiplication, the first p entries of the last column of E_k. The matrix obtained is $E_k\Pi = \begin{bmatrix} \widehat{B}_k & 0 \\ \mathbf{z}_k^{\mathrm{T}} & b \end{bmatrix}$, where $\widehat{B}_k\mathbf{z}_k = \mathbf{y}_k$, and $\mathbf{z}_k$ is formed of uncorrelated components.

The adaptive pre-whitening strategy of [56] consists of a linear prediction approach to the estimation of lower-triangular Cholesky factor of R_y^{-1}. This task reduces to the adaptive solution of $(q-1)$ linear prediction problems and the estimation of the prediction-error variances.

In [95] decorrelation is accomplished through a self-adaptive procedure of the Gram-Charlier type, looking for a lower-triangular Cholesky factor of R_y^{-1}, $\widehat{B}_k^{\dagger} = T_k = \big(t_{ij}(k)\big)$, $t_{ij} = 0$, $j > i$, $t_{ii} = 1$, with:

$$z_i(k) = y_i(k) + \sum_{j \leqslant i-1} t_{ij}(k-1)y_j(k) \tag{7.4}$$

$$t_{ij}(k) = t_{ij}(k-1) - \lambda_k z_i(k)y_j(k), \quad j \leqslant i-1. \tag{7.5}$$

This procedure is reminiscent of that of [56]. Normalization is later achieved by a kind of automatic gain control. This could perhaps be further simplified by including the normalization weights in the diagonal entries of matrix T_k, updating them as: $t_{ii}(k) = t_{ii}(k-1) - \lambda_k\big(v_i^2(k) - 1\big)$, where $v_i^2(k)$ is any adaptive estimate of the whitener output variance at time instant k.

On the other hand, in [29] the on-line pre-whitening is based on the adaptive minimization of the Kullback-Leibler divergence between the whitener output and a random vector with identity covariance matrix, when their respective distributions are truncated at the second order, i.e., when a Gaussian approximation is used. By employing the relative gradient of this contrast, one arrives at the following serial updating rule for the whitening matrix:

$$\widehat{B}_{k+1}^{\dagger} = \left[I - \lambda_k(\mathbf{z}_k\mathbf{z}_k^{\mathrm{T}} - I)\right]\widehat{B}_k^{\dagger}, \tag{7.6}$$

where $\mathbf{z}_k$ is the whitener output, given by (7.2).

The analysis of these adaptive pre-whitening procedures is beyond the scope of this investigation.

7.2.2 Adaptive unitary-matrix estimation

After pre-whitening, matrix Q needs to be estimated. The on-line estimation of this matrix has previously been carried out through a number of procedures, such as those described in Section 7.1. Here, we will concentrate on the adaptive extension of the closed-form solutions considered in the previous chapters. The obtained results will then be combined with existing pre-whitening strategies, such as that in (7.6), in order to build a complete adaptive separation system.

7.3 Adaptive Centroid Estimation or Cost-Function Stochastic Optimization?

The adaptive extension of the batch procedures of the previous chapters could be accomplished in two different but related ways: 1) the direct adaptive estimation of the corresponding centroids, which later would be plugged into the respective equations for the estimation of the relevant parameter [eqns. (3.32) and (5.4)], and 2) the stochastic optimization of the associated cost functions. In this section, we discuss why the first option is the most attractive. With such purposes in mind, we restrict this preliminary study to the EML estimator only, but the conclusions are also readily extensible to the AEML.

7.3.1 Direct adaptive centroid estimation

The derivation of an adaptive EML algorithm can be reduced to the adaptive estimation of the whitened-sensor centroid location (3.21) and the sks (3.27). To simplify the notation, let ν_k and ν_k' denote the complex scatter-diagram point of the pre-whitened observations and the sources, respectively, at time instant k:

$$\left.\begin{aligned}\nu_k &= z_1(k) + jz_2(k) = \rho_k \mathrm{e}^{j\phi_k} \\ \nu'_k &= x_1(k) + jx_2(k) = \rho_k \mathrm{e}^{j\phi'_k}\end{aligned}\right\} \nu_k = \nu'_k \mathrm{e}^{j\theta}. \tag{7.7}$$

Then, equations (3.21) and (3.27) can be written, respectively, as

$$\xi = \mathrm{E}[\nu^4] \tag{7.8}$$

$$\gamma = \mathrm{E}[|\nu|^4] - 8, \tag{7.9}$$

and hence ξ and γ may be obtained in an adaptive fashion from:[3]

$$\xi_k = (1-\mu_k)\xi_{k-1} + \mu_k \nu_k^4 \tag{7.10}$$

$$\gamma_k = (1-\mu_k)\gamma_{k-1} + \mu_k(|\nu_k|^4 - 8), \tag{7.11}$$

If $\mu_k = \frac{1}{k}$, then (7.10) and (7.11) become the adaptive equivalents of the sample estimates (3.29) and (3.30), respectively. The rotation angle θ is estimated at each iteration simply by plugging the above two parameters into (3.32):

$$\hat{\theta}_k = \frac{1}{4} \arg\big(\xi_k \cdot \mathrm{sign}(\gamma_k)\big). \tag{7.12}$$

Finally, the unitary matrix $\widehat{Q}_k = \mathcal{Q}(\hat{\theta}_k)$ estimates the source signals when applied over the whitened vector, as in eqn. (7.3).

7.3.2 Cost-function stochastic optimization

Alternatively, one can accomplish the adaptive estimation by resorting to a stochastic optimization across the cost-function surface. For simplicity, let us call $J = J_{\mathrm{EML}}$ the cost function (4.14) associated to the EML estimator. Unknown parameter θ can be estimated through the iterative steepest ascent scheme:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mu_k \nabla J, \qquad \text{with } \nabla J = \frac{\partial J}{\partial \hat{\theta}}, \tag{7.13}$$

and where μ_k is the adaption coefficient at iteration k. According to (4.7), $\nabla J = 4\gamma \mathrm{Im}\big(\xi_4(\mathbf{s})\big)$. The above iterative procedure is turned into a stochastic adaptive algorithm just by dropping the expectation operator which implicitly appears in the

[3] Being consistent with the notation employed so far, we should be using "$\hat{\xi}_k$" and "$\hat{\gamma}_k$" for the adaptive estimates of ξ and γ, respectively. However, the 'hats' will be dropped in the sequel for the sake of simplicity.

expressions of $\mathbb{I}\mathrm{m}(\cdot)$ and γ:

$$\mathbb{I}\mathrm{m}\big(\xi_4(\mathbf{s})\big) = 4\mathrm{E}[s_1^3 s_2 - s_1 s_2^3] \tag{7.14}$$

$$\gamma = \mathrm{E}[(s_1^2 + s_2^2)^2 - 8] \tag{7.15}$$

Remark, however, only one of them can be dropped at a time, whereas the other parameter needs to be adaptively estimated. Effectively, if both expectations were simultaneously omitted, the updating term in (7.13) would be

$$\mu_k 16\big[(s_1^2 + s_2^2)^2 - 8\big](s_1^3 s_2 - s_1 s_2^3). \tag{7.16}$$

This would correspond to the maximization of an 8th-order function proportional to $\mu_{80}^s + \mu_{08}^s + \mu_{62}^s + \mu_{26}^s - 12(\mu_{40}^s + \mu_{04}^s)$,[4] which little or nothing has to do with our original objectives. Therefore, only one expectation term is dropped. Assume that we choose to adaptively estimate the sks γ, e.g., via (7.11). Taking into account that $s_1(k) + js_2(k) = \nu_k \mathrm{e}^{-j\hat{\theta}_k}$, the resulting stochastic update accepts the compact reformulation (constant factors are 'absorbed' in the adaption step):

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mu_k \gamma_k \mathbb{I}\mathrm{m}(\nu_k^4 \mathrm{e}^{-j4\hat{\theta}_{k-1}}). \tag{7.17}$$

7.3.3 Discussion

The procedures presented in Sections 7.3.2 and 7.3.1 are, in principle, just as valid as adaptive implementations of the EML method. However, the first one presents certain advantages lacked by the second. We illustrate this preference with an example (perhaps abusively, since their convergence properties and asymptotic performance have not been studied yet; this will be done in Section 7.4).

Consider an orthogonal mixture of two uniformly distributed sources, with $\theta = 30^{\mathrm{o}}$. Figure 7.2 shows the trajectory, with respect to time, of the angle estimated by three procedures with $\mu = 10^{-3}$. The solid line corresponds to method (7.12), the dashed line to (7.17), and the dash-dotted line also to (7.17) but using an averaged centroid estimate [like in (7.10)] in place of the instantaneous value ν_k. It is clear that direct centroid estimation method (7.12) provides the fastest convergence and a very good accuracy.

[4]This function is easily obtained by integration of (7.16) plus later replacement of the mathematical expectations.

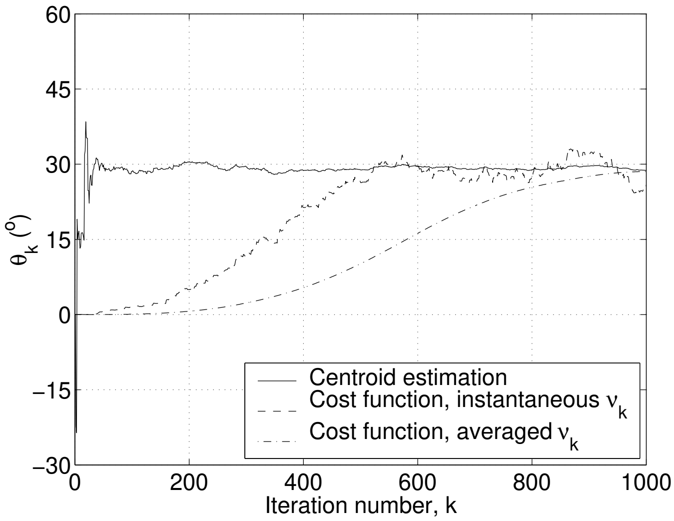


Figure 7.2: Adaptive angle estimation for two uniform sources, $\theta = 30^o$, $\mu = 10^{-3}$. Solid line: via (7.12). Dashed line: via (7.17). Dash-dotted line: via (7.17) with averaged ν_k.

Optimization-based method (7.17) converges far more slowly, and its accuracy is much poorer than the latter. The third method gives the smoothest results, but at the expense of a very slow convergence. Figure 7.3 shows the results for a new step size, $\mu = 5 \cdot 10^{-3}$. Now, the last two methods converge much faster, sacrificing precision, as commonly occurs for adaptive algorithms. But surprisingly, the first method shows an almost identical trajectory to that for the previous adaption coefficient, which was 5 times smaller.

This baffling outcome may be accounted for as follows. In order for the first method to estimate the missing angle, the magnitude of relevance is not the exact centroid location itself, but its orientation. This orientation is estimated with great accuracy in just a few iterations when the centroid is initialized at the origin of the complex plane ($\xi_0 = 0$). As an illustration, Figure 7.4 plots the centroid trajectory[5] for another mixture of two uniform sources, with a true angle of 30^o [providing a centroid orientation of $\angle\xi = 4 \times 30^o + 180^o = 300^o = -60^o$, according to (3.24)] and adaption parameter $\mu = 2 \cdot 10^{-3}$. Further experiments (see Figure 7.5) also confirm that the convergence

[5]The 'ODE solution' (dashed line) will be mathematically introduced in Section 7.4.2. Basically, it represents a theoretical approximation to the empirical trajectory.

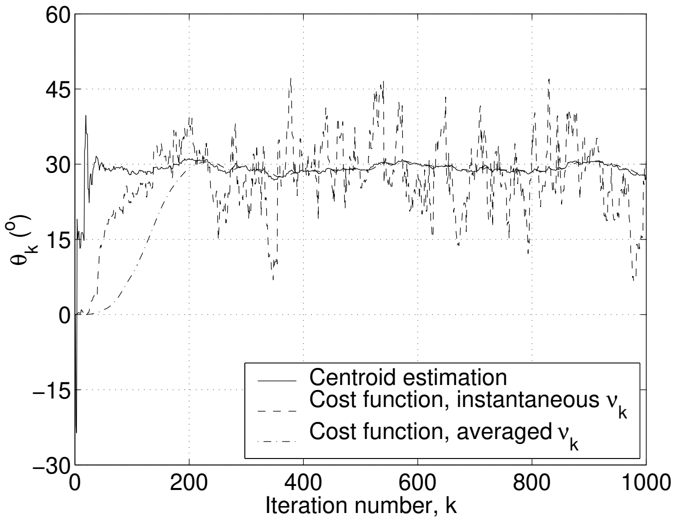


Figure 7.3: Adaptive angle estimation for two uniform sources, $\theta = 30^o$, $\mu = 5 \cdot 10^{-3}$. Solid line: via (7.12). Dashed line: via (7.17). Dash-dotted line: via (7.17) with averaged ν_k.

speed is practically independent of the particular choice of the adaption coefficient.

In short, we prefer the adaptive centroid estimation procedure relative to those based on cost-function optimization for various reasons: 1) excellent balance between convergence speed and accuracy, 2) high convergence speed, extremely robust with respect to the adaption coefficient, and 3) for the adaptive use of the combined estimation strategy introduced in Chapter 5, the centroids need to be calculated regardless. On these grounds, we focus our efforts into studying the adaptive procedure (7.12), the results for (7.17) being obtained in a very similar manner. Estimator (7.12) is referred to as *adaptive EML (adEML)*. An exhaustive analysis of its convergence and asymptotic properties is carried out next.

7.4 Asymptotic Analysis

7.4.1 Preliminary results

In the following, and for the sake of simplicity, a constant adaption coefficient will be assumed, $\mu_k = \mu$, $\forall k$. This choice, in turn, allows the tracking of non-stationarities in the mixing system, which is one of the main rationales behind the use of adaptive

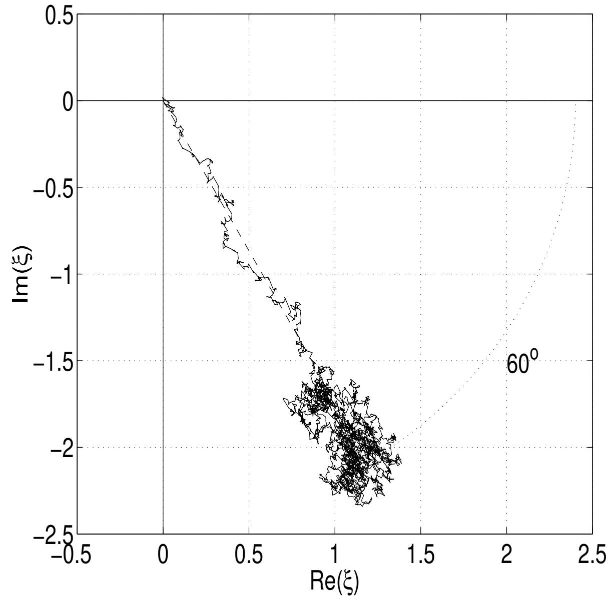


Figure 7.4: An example of the centroid trajectory for a mixture of two uniform distributions, $\mu = 2 \cdot 10^{-3}$, first $5 \cdot 10^3$ iterations. Solid line: adaptive estimator (7.10). Dashed line: ODE solution (7.27).

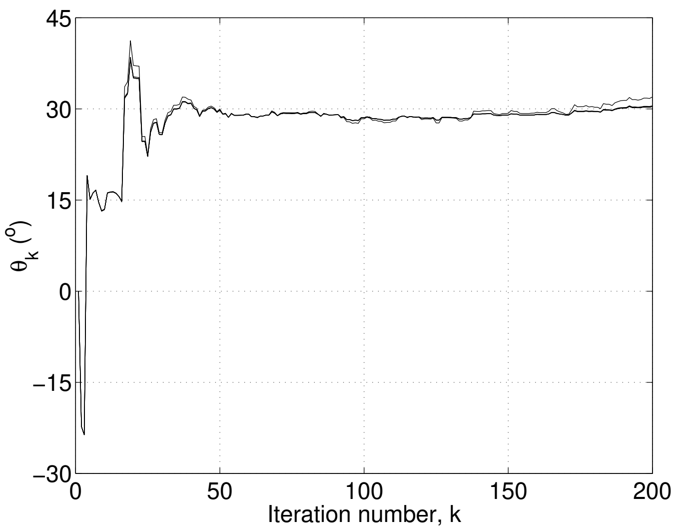


Figure 7.5: Superimposed angle trajectories obtained through (7.12) for adaption coefficients $\mu = \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$ and a fixed mixture realization of uniformly-distributed sources. Convergence is invariably reached in all cases at around the 50th sample.

separation procedures. In that case, (7.10) may be written in the form:

$$\xi_k = \mu \sum_{n=1}^{k} (1-\mu)^{k-n} \nu_n^4. \tag{7.18}$$

Taking expectations at both sides of the previous equation, and assuming i.i.d. mixtures:

$$\mathrm{E}[\xi_k] = [1-(1-\mu)^k]\,\mathrm{E}[\nu^4] = [1-(1-\mu)^k]\,\xi \xrightarrow[k\to\infty]{} \xi, \qquad 0<\mu<2. \tag{7.19}$$

That is, adaptive estimator (7.10) of centroid (3.21) is *asymptotically unbiased*. On the other hand, the variance of ξ_k becomes

$$\mathrm{Var}[\xi_k] = \frac{\mu}{2-\mu}[1-(1-\mu)^{2k}]\,\mathrm{Var}[\nu^4] \xrightarrow[k\to\infty]{} \frac{\mu}{2-\mu}\mathrm{Var}[\nu^4], \tag{7.20}$$

so that it can be made arbitrarily small by choosing μ sufficiently small. Totally analogous conclusions hold for adaptive estimator (7.11) of the sks. These preliminary asymptotic results lead to

$$\mathrm{E}[\hat{\theta}_k] \xrightarrow[k\to\infty]{} \theta \tag{7.21}$$

if the sks is not null, since in that case $\hat{\theta}_{\mathrm{EML}}$ given by (3.32) is also asymptotically unbiased (Section 3.3.5).

7.4.2 In-depth analysis

Adaption scheme (7.10)–(7.11) is appealing in that it accepts the typical adaptive algorithm form [11]:

$$\xi_k = \xi_{k-1} + \mu_k H(\xi_{k-1}, \nu_k), \tag{7.22}$$

where, from (7.10), function $H(\cdot,\cdot)$ can be identified as

$$H(\xi, \nu) = \nu - \xi, \tag{7.23}$$

and accordingly for (7.11). As a result, the standard tools for the study of stochastic algorithms of this general form can be employed to analyze the convergence properties of the one suggested herein. Our main goal is to obtain the asymptotic pdf of the adEML estimator (7.12).

The ODE method

In the first place, the ordinary differential equation (ODE) method [4,11,56] is employed to study the trajectories of the centroid locations (7.22). This method approximates the true trajectories of the parameters of interest by the discretized solution of certain ODE. It is shown [11] that in the limit of an arbitrarily small adaption coefficient, the true trajectories followed by ξ_k converge to the continuous function of time $\xi(t)$ discretized at time instants $t_k = k\mu$, $\xi(t)$ being given by the solution of the ODE:

$$\dot{\xi}(t) = h(\xi). \tag{7.24}$$

The term $h(\cdot)$, called *mean field*, is obtained from the function $H(\cdot, \cdot)$ as:

$$h(\xi) \triangleq \mathrm{E}[H(\xi,\, \nu_k)] = \xi_* - \xi, \tag{7.25}$$

where

$$\xi_* = \mathrm{E}[\nu_k], \tag{7.26}$$

is the population centroid location (3.21) [or (7.8)], and symbol $\mathrm{E}[\cdot]$ denotes the expectation with respect to the distribution of the state ν_k for a fixed value of ξ when ν_k is asymptotically stationary. The conditions for the validity of this approximation are, essentially, that the adaptation coefficient μ be very small and the function $\xi \mapsto H(\xi,\, \nu)$ be fairly regular. Having a look at (7.23), it is clear that the latter condition holds. In our case, it is straightforward to obtain

$$\xi(t) = \xi_* + (\xi_0 - \xi_*)\,\mathrm{e}^{-t}, \tag{7.27}$$

where ξ_0 is the initial value, i.e., $\xi_0 = \xi(0)$. For the source kurtosis trajectories the solution of its ODE is also readily found as:

$$\gamma(t) = \gamma_* + (\gamma_0 - \gamma_*)\,\mathrm{e}^{-t}, \tag{7.28}$$

where, again, $\gamma_0 = \gamma(0)$ is the initial value and

$$\gamma_* = \mathrm{E}[|\nu_k| - 8] \tag{7.29}$$

represents the population sks (3.27) [or (7.9)].

Stability of the equilibrium points

The *equilibrium points* or *attractors* of (7.22) are those values of ξ such that the mean field $h(\cdot)$ vanishes. Therefore, $\xi = \xi_*$ is the unique attractor of (7.10) [or (7.22)], as seen from (7.25) and taking $\lim_{t \to \infty} \xi(t)$ in (7.27). An attractor is said to be *locally asymptotically stable (LAS)* if and only if [11]:

$$h_\xi(\xi_*) \triangleq \left. \frac{\partial h}{\partial \xi} \right|_{\xi = \xi_*} \tag{7.30}$$

has a negative real part. In the case of interest, and according to (7.25)

$$h_\xi(\xi_*) = -1 < 0, \tag{7.31}$$

and hence the equilibrium point ξ_* is LAS always, regardless of the source distribution and the mixing structure. Analogously, $\gamma = \gamma_*$ is the unique LAS equilibrium point of its respective ODE. Therefore, the attractors of (7.12) are LAS too. On the other hand, since ξ_k and γ_k converge to their population values, the adEML equilibrium points constitute valid separation solutions, under the EML basic condition on the sks. In conclusion, under such condition, the attractors of adaptive estimator (7.12) provide LAS valid separation solutions.

Asymptotic normality of the centroid locations

Under the above conditions, it is proved in [11, Theorem 2, Chapter 3] that the asymptotic distribution of ξ_k is normal. Specifically:

$$\tilde{\xi_k} \triangleq \mu^{-\frac{1}{2}} \big(\xi_k - \xi(t_k)\big) \xrightarrow[\substack{k \to \infty \\ \mu \to 0}]{} \mathrm{N}(0,\, P), \tag{7.32}$$

where the covariance $P \triangleq \mathrm{Cov}[\tilde{\xi_k}]$ is given by the solution of the Lyapunov equation:

$$2 h_\xi(\xi_*) P + R(\xi_*) = 0, \tag{7.33}$$

with

$$R(\xi) \triangleq \sum_{k=-\infty}^{\infty} \mathrm{Cov}\big[H(\xi,\, \nu_k),\, H(\xi,\, \nu_0)\big]. \tag{7.34}$$

In conclusion, the distribution of the centroid ξ_k is asymptotically normal

$$\xi_k \xrightarrow[\substack{k\to\infty \\ \mu\to 0}]{} \mathrm{N}(\xi_*, \sigma_\xi^2), \tag{7.35}$$

with mean ξ_* and variance

$$\sigma_\xi^2 = \mu P = \mu R(\xi_*)/2. \tag{7.36}$$

The pdf of ξ_k for k large enough and μ sufficiently small is then:

$$p_\xi(\xi) = \frac{1}{2\pi\sigma_\xi^2} \exp\left(-\frac{\|\xi - \xi_*\|^2}{2\sigma_\xi^2} \right), \tag{7.37}$$

which, as a function of the real and imaginary part of $\xi = \omega_1 + j\omega_2$, becomes the product of two uncorrelated real Gaussian pdf's:

$$p_\xi(\omega_1, \omega_2) = \frac{1}{2\pi\sigma_\xi^2} \exp\left(\frac{-(\omega_1 - \omega_{1_*})^2 - (\omega_2 - \omega_{2_*})^2}{2\sigma_\xi^2} \right), \qquad \xi_* = \omega_{1_*} + j\omega_{2_*}. \tag{7.38}$$

Now we are ready to develop an expression for the asymptotic pdf of the adaptive angle estimator (7.12).

Asymptotic pdf

First, if k is sufficiently large and μ sufficiently small, it can be reasonably assumed that the sign of the sks is accurately estimated, so that it becomes a constant and then the "sign(·)" function in (7.12) may be neglected. In those conditions, and assuming with no loss of generality that such a sign is positive, we have:

$$\delta \triangleq 4\hat{\theta} = \arg(\xi). \tag{7.39}$$

This last expression naturally induces the change of variables:

$$\begin{cases} \omega_1 = r\cos\delta \\ \omega_2 = r\sin\delta, \end{cases} \tag{7.40}$$

and integrating with respect to r, the marginal pdf of δ can be obtained as:

$$p_\delta(\delta) = \int_0^\infty \frac{r}{2\pi\sigma_\xi^2} \exp\left(\frac{-(r\cos\delta - \omega_{1_*})^2 - (r\sin\delta - \omega_{2_*})^2}{2\sigma_\xi^2} \right) dr. \tag{7.41}$$

With the help of the symbolic mathematics package MAPLE™ [36] and after some tedious algebraic simplifications, p_δ turns out to be:

$$p_\delta(\delta) = \frac{1}{2\pi}\exp\left(\frac{-\gamma^{+2}}{2\sigma_\xi^2}\right) +$$
$$+\frac{1}{\sigma_\xi\sqrt{8\pi}}\gamma^+\cos(\delta-4\theta)\exp\left(\frac{-\gamma^{+2}\sin^2(\delta-4\theta)}{2\sigma_\xi^2}\right)\left\{1+\operatorname{erf}\left(\frac{1}{\sigma_\xi\sqrt{2}}\gamma^+\cos(\delta-4\theta)\right)\right\}, \tag{7.42}$$

where the symbol γ^+ denotes the absolute value of the sks:

$$\gamma^+ \triangleq |\gamma_*| = |\kappa_{40}^x + \kappa_{04}^x|. \tag{7.43}$$

With γ^+ defined as above, it is easy to prove that pdf (7.42) is also applicable to the case of negative sks. Finally, from (7.39):

$$p_{\hat{\theta}}(\hat{\theta}) = 4\,p_\delta(4\hat{\theta}), \tag{7.44}$$

which is the pdf of $\hat{\theta}_k$ given by (7.12) when $k \to \infty$ and $\mu \to 0$.

Pdf $p_{\hat{\theta}}$ is an even function of $(\hat{\theta} - \theta)$, and therefore:

$$\mathrm{E}[\hat{\theta}] = \theta, \tag{7.45}$$

that is, adaptive estimator (7.12) is *asymptotically unbiased*. Moreover, the smaller the ratio σ_ξ/γ^+, the more 'peaky' the pdf (7.42) becomes. Hence, the asymptotic variance of estimator (7.12), which measures the variability of the estimates once convergence is reached, is somehow related to that ratio:

$$\mathrm{Var}[\hat{\theta}] \triangleq \sigma_{\hat{\theta}}^2 \sim \frac{\sigma_\xi^2}{\gamma^{+2}}. \tag{7.46}$$

The specific trend of this relation is obtained by means of numerical integration of $\int_{-\pi}^{\pi} \delta^2 p_\delta(\delta)\, d\delta$, with $p_\delta(\delta)$ given by (7.42). The plot of the standard deviation $\sigma_{\hat{\theta}}$ versus the ratio σ_ξ/γ^+ is represented in Figure 7.6. For $\sigma_\xi/\gamma^+ < 0.138$, $\sigma_{\hat{\theta}}$ can be approximated by the straight line

$$\sigma_{\hat{\theta}} = \frac{\sigma_\xi}{4\gamma^+} \tag{7.47}$$

with an error of less than 1%. Fig. 7.6 also displays this linear approximation. Note also that, from (7.42) and (7.44), when the ratio σ_ξ/γ^+ becomes large (which occurs

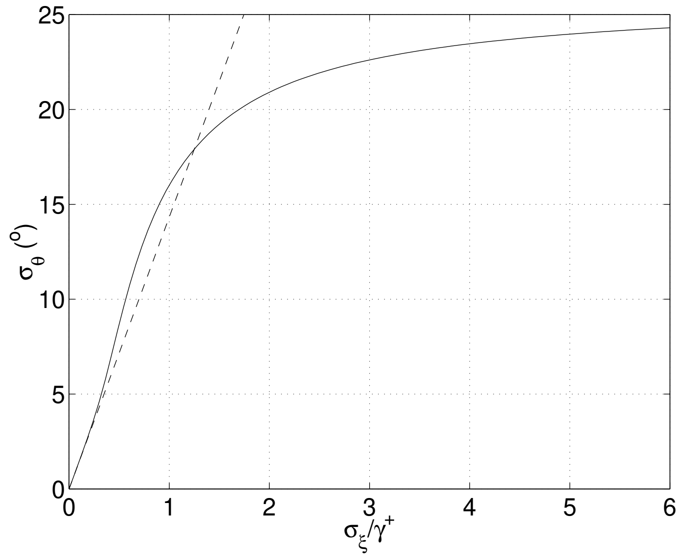


Figure 7.6: Asymptotic standard deviation of the adEML estimator, $\sigma_{\hat{\theta}}$, versus σ_ξ/γ^+. Solid line: obtained from analytical pdf (7.42). Dashed line: linear approximation (7.47) valid for small σ_ξ/γ^+.

when either the step size is too large [eqn. (7.36)] or the sks is close to zero), the distribution of $\hat{\theta}_k$ tends to a uniform pdf $\mathrm{U}(-\frac{\pi}{4}, \frac{\pi}{4})$, with standard deviation 25.98°. This result agrees with the behaviour observed in the solid line of Fig. 7.6. Since $\sigma_\xi^2 \propto \mu$ [eqn. (7.36)], it follows that the variance $\sigma_{\hat{\theta}}^2$ can be made as small as desired just by choosing the appropriate value of the adaption coefficient μ, at the expense of slowing down the convergence speed. Eqn. (7.47) is theoretically substantiated next.

Gaussian approximation

Similarly as occurred for the batch EML method (Section 3.3.5), asymptotic pdf (7.42) admits a Gaussian approximation. If μ is small enough, the ratio σ_ξ/γ^+ becomes small too [because $\sigma_\xi \propto \sqrt{\mu}$, eqn. (7.36)], and hence $\exp\bigl(-{\gamma^+}^2/(2\sigma_\xi^2)\bigr) \approx 0$. In addition, the pdf is mostly concentrated around $\delta = 4\theta$, so that $\cos(\delta - 4\theta) \approx 1$, $\sin(\delta - 4\theta) \approx (\delta - 4\theta)$ and $\mathrm{erf}(\cdot) \approx 1$. Consequently, pdf (7.42) reduces to:

$$p_\delta(\delta) \underset{\mu\to 0}{\approx} \hat{p}_\delta(\delta) = \frac{1}{\sqrt{2\pi}\,\sigma_\xi/\gamma^+} \exp\left\{ \frac{-(\delta - 4\theta)^2}{2(\sigma_\xi/\gamma^+)^2} \right\}, \tag{7.48}$$

that is:

$$\delta \xrightarrow[\mu \to 0]{\mathrm{d}} \mathrm{N}\big(4\theta,\, (\sigma_\xi/\gamma^+)^2\big). \tag{7.49}$$

From this result and relation (7.39), the asymptotic variance of the adEML is readily obtained as:

$$\mathrm{Var}[\hat{\theta}_k] \xrightarrow[\substack{k\to\infty \\ \mu\to 0}]{} \sigma^2_{\text{adEML}} = \frac{\sigma_\xi^2}{16(\kappa_{40}^x + \kappa_{04}^x)^2}. \tag{7.50}$$

This matches the estimate empirically anticipated in eqn. (7.47).

Remarks

In order to bring this analysis to an end, there are a couple of remarks worth pointing out. In the first place, note that from equations (3.12)–(3.13), (3.24) and (7.7), function $H(\cdot,\,\cdot)$ in (7.23) can be written as a function of the source signals and the angle θ:

$$H(\xi_*,\, \nu_k) = \nu_k^4 - \xi_* = \big[\big(x_1(k) + jx_2(k)\big)^4 - (\kappa_{40}^x + \kappa_{04}^x)\big]\mathrm{e}^{j4\theta} = (\nu_k'^{\,4} - \gamma)\mathrm{e}^{j4\theta}, \tag{7.51}$$

and then expression (7.34) becomes:

$$R(\xi_*) = \sum_{k=-\infty}^{\infty} \mathrm{E}\big[(\nu_k^4 - \xi_*)(\nu_0^4 - \xi_*)^*\big] = \sum_{k=-\infty}^{\infty} \mathrm{E}\big[(\nu_k'^{\,4} - \gamma_*)(\nu_0'^{\,4} - \gamma_*)^*\big]. \tag{7.52}$$

The complex conjugation makes θ vanish in the above equation, so that R does not depend on the whitened observations and the parameter of interest θ, but exclusively on the source signals. Assuming i.i.d. processes, R reduces to

$$R(\xi_*) = \mathrm{Var}[\nu'^4] = \mathrm{Var}[(x_1 + jx_2)^4] = \mathrm{Var}[\tilde{\omega}_1 + j\tilde{\omega}_2] = \tilde{\sigma}_1^2 + \tilde{\sigma}_2^2, \tag{7.53}$$

where $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2$ are given by (3.55) and (3.57), respectively. Hence, connecting eqns. (7.36), (7.50) and (7.53), the asymptotic variance of the adEML finally reads:

$$\sigma^2_{\text{adEML}} = \frac{\mu(\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2)}{32(\kappa_{40}^x + \kappa_{04}^x)^2}. \tag{7.54}$$

Observe also that in the case of i.i.d. sources, the variance σ_ξ^2 obtained from this value of R coincides with the result for this kind of sources initially given in (7.20) when $\mu \to 0$. More importantly, from the above expression it turns out that the variance, and hence the performance, of the adEML estimator does not depend on the mixture, but only on the source signals, and more specifically, on their statistics up to the eighth order. The uniform performance property of batch estimator (3.32), proven at the end of Section 3.3.5, is therefore inherited by its adaptive counterpart (7.12).

Finally, it is also interesting to realize the striking resemblance between the asymptotic pdf of the adaptive estimator (7.12) and the large-sample asymptotic pdf found in Section 3.3.5 for its block-processing version (3.32): both are identical when $\sigma_1 = \sigma_2 = \sigma_\xi$.

7.5 Extension to General BSS Scenario

The extension of adaptive scheme (7.12) to the broader BSS scenario in which more than two whitened observations appear is very similar to that for its batch counterpart described in Section 6.4.1 and Fig. 6.2. Now, at each time instant a whitened-signal sample is first obtained through the suitable adaptive pre-whitening method. Then, the resulting vector sample is processed in sweeps over all its component pairs. For each sample pair, the associated centroid and sks estimates are updated, as in eqns. (7.10)–(7.11), and the resulting angle estimate (7.12) obtained. From it, a counter-rotation aiming at separation is performed on the sample pair. The process is repeated in turn over all the other pairs, and over several sweeps. It is important to remark that the values of the centroid and sks associated with a given signal pair are independently stored for each sweep. If the signals are stationary, the angle values obtained by this adaptive procedure will converge to those of its batch counterpart. This pairwise extension, which is outlined in Figure 7.7, is very similar to that suggested in [41] for the AROT method. The two adaptive procedures described next can also be extended to more than two signals in a totally analogous manner.

7.6 Adaptive AEML and Combined Estimation

We now discuss the adaptive implementation of the AEML and the combEML methods developed in batch form in Chapter 5. Most results exposed in Sections 7.3–7.5 are easily generalized to those two estimators, and hence the specific developments are omitted.

At each sample instant k:

Step 1. Obtain whitened-output sample from current observation sample via a suitable adaptive pre-whitening method. Consider the resulting whitening-matrix (pseudo)inverse and the whitened sample as a first estimate of the mixing matrix and the source sample, respectively: $\widehat{B}_k \to \widehat{M}_k$, $\mathbf{z}_k \to \hat{\mathbf{x}}_k$.

Step 2. For each sweep number c repeat Step 3:

Step 3. For all signal pairs $\hat{\mathbf{x}}_{ij} = [\hat{x}_i, \hat{x}_j]^{\mathrm{T}}$:

- 3.1. Update centroid $\xi_{ij}^{(c)}(k)$ and sks $\gamma_{ij}^{(c)}(k)$ using eqns. (7.10) and (7.11) (with $[z_1, z_2]^{\mathrm{T}} = [\hat{x}_i, \hat{x}_j]^{\mathrm{T}}$).
- 3.2. Estimate (2×2)-unitary transformation $\widehat{Q}_{ij}^{(c)}$ from adEML estimator (7.12).
- 3.3. Counter-rotate: update $\hat{\mathbf{x}}_{ij} := \big(\widehat{Q}_{ij}^{(c)}\big)^{\mathrm{T}} \hat{\mathbf{x}}_{ij}$, $\widehat{M}_{k:[i,j]} := \widehat{M}_{k:[i,j]} \widehat{Q}_{ij}^{(c)}$, where $\widehat{M}_{k:[i,j]}$ denotes the matrix composed of only the ith and jth columns of $\widehat{M}_k$.

Figure 7.7: Adaptive algorithm for BSS of more than two signals via pairwise processing with the adEML estimator.

7.6.1 Adaptive AEML estimator

As for the EML, the adaptive extension of the AEML can also be made according to the two strategies exposed in Section 7.3. However, for the same reasons given at the end of that section, the implementation based on the direct estimation of the corresponding centroid is preferred. From (5.5) and (7.7), this updating reads:

$$\xi'_k = (1 - \mu_k)\xi'_{k-1} + \mu_k |\nu_k|^2 \nu_k^2. \tag{7.55}$$

The angle estimation is simply accomplished by plugging the value of ξ'_k into (5.4), yielding:

$$\hat{\theta}_k = \frac{1}{2} \arg(\xi'_k). \tag{7.56}$$

We call eqn. (7.56) the *adaptive AEML (adAEML)* estimator.

The asymptotic analysis of this algorithm is carried out along the lines of the analysis of Section 7.4. We therefore skip the details. Essentially, the LAS equilibrium points of (7.56) are all separating solutions. Its asymptotic pdf corresponds to (7.42), but with the parameter substitutions: $\gamma^+ \to \eta^+ \triangleq |\eta| = |\kappa_{40}^x - \kappa_{04}^x|$, and $\sigma_\xi = \tilde{\sigma}_1'^2 + \tilde{\sigma}_2'^2$, where

$\tilde{\sigma}_1'^2$ and $\tilde{\sigma}_2'^2$ are the variances obtained from (5.18), and accordingly for its Gaussian approximation (7.48). With these minor differences in mind, the asymptotic variance of the adAEML can be expressed as:

$$\mathrm{Var}[\hat{\theta}_k] \xrightarrow[\substack{k\to\infty \\ \mu\to 0}]{} \sigma^2_{\mathrm{adAEML}} = \frac{\mu(\tilde{\sigma}_1'^2 + \tilde{\sigma}_2'^2)}{8(\kappa_{40}^x - \kappa_{04}^x)^2}. \tag{7.57}$$

7.6.2 Adaptive combined EML estimator

Due to the analogy between the asymptotic variances of the batch methods [eqns. (3.69) and (5.19)] and those of the adaptive procedures [eqns. (7.54) and (7.57)], and according to the exposition of Section 5.3, practical decision rule (5.42) appears as a reasonable strategy for the selection between the adEML and the adAEML in order to derive an analogous adaptive combined estimation scheme. Further, the explicit estimation of the centroids (7.10) and (7.55) in adaptive procedures adEML (7.12) and adAEML (7.56), respectively, makes straightforward the adaptive extension of such decision principle. In adaptive fashion, the rule is evaluated at each iteration k as:

$$|\xi_k| \underset{\mathrm{adAEML}}{\overset{\mathrm{adEML}}{\gtrless}} |\xi_k'|, \tag{7.58}$$

and the selected estimate, adEML or adAEML, is applied accordingly. This simple extension gives rise to the *adaptive combined EML (adcombEML)* procedure, and is evaluated in some of the experiments included in Section 7.8.

7.7 Computational Complexity

We now measure the computational burden of the adaptive schemes presented in the previous sections. Since additions are negligible relative to products, we consider a flop as a real multiplication. Recall (Section 2.4.5) that a complex product takes 4 flops, and the multiplication of a real by a complex quantity (real-complex product) 2 flops. At each time instant (iteration) k and for each signal pair $[\hat{x}_i, \hat{x}_j]^{\mathrm{T}}$, the adEML source-extraction algorithm described in Section 7.5 and Fig. 7.7 can be broken down in the following elementary steps:

1) Computation of ν_k^4 (7.7): 2 complex products = 8 flops. The computation of $|\nu_k^4|$

takes no additional flops, since it can be evaluated as $|\nu_k^2|^2$ using interim results in the calculation of ν_k^4.

2) Centroid updating (7.10): 2 real-complex products = 4 flops. Sks updating (7.11): 2 real products = 2 flops.

3) Computation of $\hat{\theta}_k$ (7.12) and rotation matrix set-up (3.2): this would be platform-dependent, as it could be done through a look-up table, or even with fast built-in functions. Say it takes f flops.

4) Rotation: 4 real products = 4 flops.

This makes $(18 + f)$ flops per signal pair per iteration. With N sweeps over the $q(q-1)/2$ signal pairs, the overall complexity per iteration of the adEML procedure becomes: $C_{\text{adEML}} = (18 + f)Nq(q-1)/2$ flop/iteration. Usually, $N = \mathrm{O}(\sqrt{q})$ (Section 6.4.1), yielding $C_{\text{adEML}} = \mathrm{O}(q^{5/2})$. For the adAEML method, the cost of centroid updating (7.55) would be 6 flops, including the computation of the term $|\nu_k|^2\nu_k^2$, so that $C_{\text{adAEML}} = (10 + f)Nq(q-1)/2 = \mathrm{O}(q^{5/2})$. Hence, the adcombEML burden would lie between C_{adEML} and C_{adAEML}, and only an extra 4 flops per signal-pair due to the modulus computation in decision rule (7.58) would have to be added.[6]

These figures can be compared with the values of other adaptive methods, such as AROT [41] and EASI [29]. The former procedure entails a higher complexity since it involves divisions[7] and square roots. However, with the equivalent centroid-based formalism proposed in Section 3.4.2 for the CF estimator, one obtains $C_{\text{AROT}} = (14+f)\cdot \cdot Nq(q-1)/2$. AROT's computational cost is thus very similar to adEML-adAEML's. By analogous reasoning, the complexity of EASI's orthogonal-matrix serial update plus source-sample extraction can be calculated as $C_{\text{EASI}} = q^3 + 3q^2 + lq = \mathrm{O}(q^3)$, where each non-linearity element takes l flops (e.g., for cubic non-linearities: $l = 2$). An extra overhead would have to be added in the normalized version of the algorithm. Typically, hence, EASI is more involved than the other methods considered above.

[6] This would be just 3 flops if the alternative decision rule derived in Chapter 6 is used (where the sks estimate is employed instead of the centroid estimate).

[7] Divisions by powers of two are neglected, since they can be efficiently accomplished by simple register shifting.

7.8 Experimental Results

Salient features of the theoretical results presented earlier in this chapter are illustrated in this section through a variety of simulations. Firstly, the theoretical asymptotic results of Section 7.4 are validated. A comparison with other adaptive angle estimators is then established. Later, it is tested how the adEML and adcombEML estimators respond to non-stationary parameters and abrupt changes of the mixing system. Finally, the performance of a whole separation system comprising the pre-whitening stage is also demonstrated and contrasted, in scenarios of two and more than two signals.

7.8.1 Asymptotic performance

Quality of ODE approximation

In the first place, it is seen how the ODE solution is indeed a good approximation of the adaptive-system actual evolution. From i.i.d. samples of two independent uniformly distributed sources and an arbitrary fixed value for the rotation angle θ, the solid line represented in Figure 7.8 shows the trajectory of the sks adaptive estimate (7.11) for $\mu = 10^{-3}$. The dashed line corresponds to the ODE solution (7.28), discretized according to $t_k = \mu k$, k being the iteration (or sample) number. Both curves are remarkably similar. The estimate settles down, roughly, from iteration $k = 5 \cdot 10^3$, corresponding to $t_k = 5$, for which the exponential function in (7.28) has dropped below 1% of its initial value.

Asymptotic-pdf validation

In the second place, the accuracy of the asymptotic pdf found in Section 7.4, formula (7.42) and its Gaussian approximation eqn. (7.48), is examined. We compare the empirical with the expected analytic results for the pdf $p_{\Delta\hat{\theta}}$ of the adEML error $\Delta\hat{\theta} \triangleq \hat{\theta} - \theta$, where $\hat{\theta}$ is given by (7.12) and $p_{\Delta\hat{\theta}}(\Delta\hat{\theta}) = p_{\hat{\theta}}(\Delta\hat{\theta} + \theta)$. Again, two independent uniformly distributed signals act as sources, with a fixed rotation angle of $\theta = 30^{\text{o}}$, yielding a hypothetical set of two whitened observations. From 10^5 independent samples of such processes, the value of $R(\xi_*)$ is evaluated by means of the sample estimate of (7.53) (it was verified that identical results are obtained when computed from the

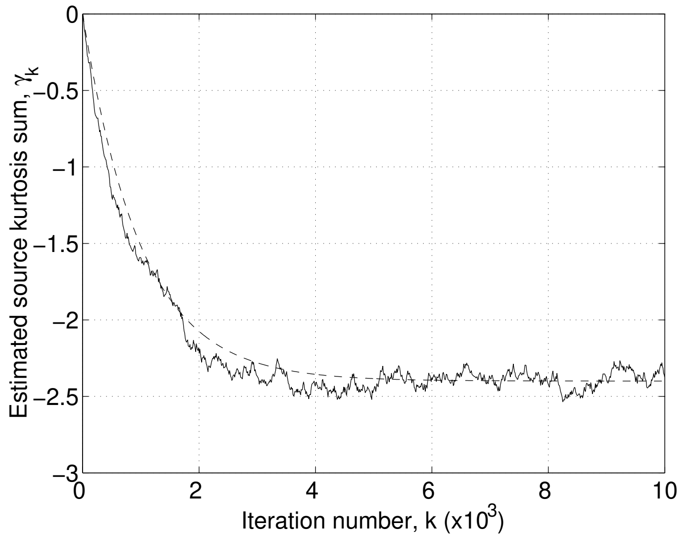


Figure 7.8: Trajectory of the adaptive source kurtosis sum estimate for a mixture of two uniformly distributed signals, $\mu = 10^{-3}$. Solid line: estimator (7.11). Dashed line: ODE approximation (7.28).

sources or the whitened signals), giving $R = 61.943$ (population value 62.537). The true value of sks gives $\gamma^+ = |-1.2 - 1.2| = 2.4$. With these parameters and $\mu = 10^{-3}$, the centroid standard deviation expected from the theoretical results [eqn. (7.36)] is $\sigma_\xi = \sqrt{\mu R/2} = 0.176$, providing a ratio $\sigma_\xi/\gamma^+ = 0.073$. Now, adaptive estimator (7.12) is used on the first 10^4 samples of the whitened observations. In the previous paragraph it was explained how the ODE solution stabilizes at $\mu k \approx 5$. This remark leads us to consider the values of $\hat{\theta}_k$ from $k = 5/\mu = 5 \cdot 10^3$, i.e., the last $5 \cdot 10^3$ iterations, in order to estimate the adEML-error empirical pdf. To this end, the kernel pdf estimation method is employed [115] with function width $w = 0.909^{\circ}$. This empirical pdf is displayed in the solid-line curve of Fig. 7.9(a). The dotted line represents the analytical pdf obtained from (7.42) and the parameters γ^+ and σ_ξ above. The dashed line corresponds to Gaussian approximation (7.48) obtained from the population values. All three curves are very similar.

Next the experiment is repeated with a new adaption coefficient $\mu = 10^{-4}$ and processing the whole of the initial 10^5 source samples. Now, $\sigma_\xi = 0.056$ and $\sigma_\xi/\gamma^+ = 0.023$. From the last $5 \cdot 10^4$ iterations, the pdf of $\hat{\theta}_k$ is estimated using again the

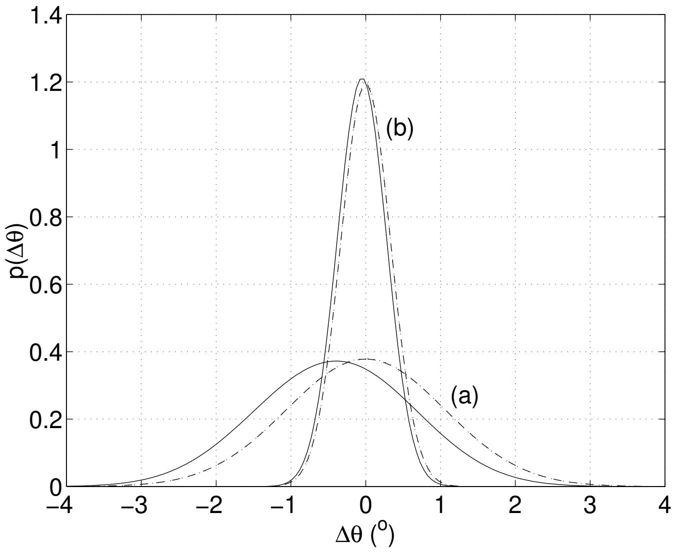


Figure 7.9: adEML-error pdf for a mixture of two independent uniform distributions, $\theta = 30^{\circ}$. Solid lines: empirical pdf estimates, kernel method with function width w. Dotted lines: analytic asymptotic pdf (7.42). Dashed lines: Gaussian approximation (7.48). (a) $\mu = 10^{-3}$, last $5 \cdot 10^3$ iterations, $w = 0.909^{\circ}$. (b) $\mu = 10^{-4}$, last $5 \cdot 10^4$ iterations, $w = 0.303^{\circ}$.

kernel method, this time with $w = 0.303^{\circ}$. This estimate appears in the solid line of Fig. 7.9(b). Formulae (7.42) and (7.48) produce the analytic curves displayed by the dotted and dashed line, respectively, in the same figure. Now the curves are even closer than before, evidencing the fact that the asymptotic results become more accurate as parameter μ gets closer to zero. As expected, the variance is reduced relative to the previous case, roughly in the same proportion as σ_{ξ}^2 decreases. This corroborates the asymptotic-variance expression (7.50).

In order to demonstrate that all these asymptotic results also hold for long-tailed and asymmetric source distributions, analogous experiments are carried out with a Laplacian (long-tailed symmetric) and an exponential (long-tailed asymmetric) distribution for the sources. The results appear in Fig. 7.10, for $\theta = 30^{\circ}$, $R = 1.771 \cdot 10^4$ (calculated from a 10^5-sample i.i.d. source realization; population value $1.902 \cdot 10^4$), $\gamma^+ = 9$ (from the population sks), and the same general conditions as above. The two cases are: (a) $\mu = 10^{-3}$, $\sigma_{\xi} = 2.976$, $\sigma_{\xi}/\gamma^+ = 0.331$, and $w = 4.444^{\circ}$; (b) $\mu = 10^{-4}$, $\sigma_{\xi} = 0.941$, $\sigma_{\xi}/\gamma^+ = 0.105$, and $w = 1.379^{\circ}$. The resemblance between the curves is

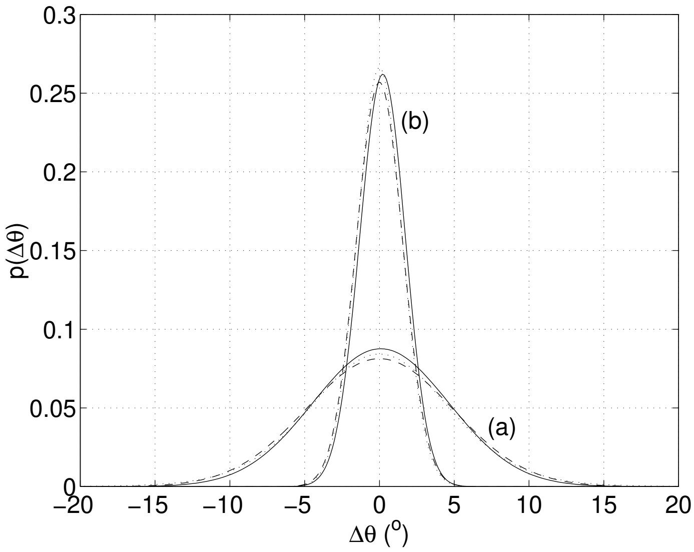


Figure 7.10: adEML-error pdf for a mixture of a Laplacian and an exponential distribution, $\theta = 30^{\text{o}}$. Solid lines: empirical pdf estimates, kernel method with function width w. Dotted lines: analytic asymptotic pdf (7.42). Dashed lines: Gaussian approximation (7.48). (a) $\mu = 10^{-3}$, last $5 \cdot 10^3$ iterations, $w = 4.444^{\text{o}}$. (b) $\mu = 10^{-4}$, last $5 \cdot 10^4$ iterations, $w = 1.379^{\text{o}}$.

also outstanding.

Remark that in all the previous examples the ratio $\frac{\sigma_\xi/\gamma^+}{w}$ is approximately constant, so that the pdf estimation can be considered as reasonably 'fair' in all cases.

7.8.2 Comparison with other adaptive angle estimators

In this section, the adEML estimator is compared with two other adaptive angle estimators derived from the EASI [29][8] and the AROT [41] methods, respectively. The pre-whitening stage in the latter procedure is removed so that only the angle estimation from the whitened observations is considered. In order to disclose the remaining orthogonal rotation, the AROT method requires to estimate the 4th-order cross-cumulants of the whitened sensor outputs, which can be done from their moments (just as suggested in [41]). To this end, the adaption coefficient λ is used to estimate the second-order

[8] Original MATLAB code was downloaded from
`http://www-sig.enst.fr/~cardoso/guidesepsou.html`.

moments, while the adaption coefficient μ is employed to estimate the fourth-order moments. For the part of the EASI method, when cubic non-linearities are chosen the convergence condition becomes that the sum of kurtosis of every pair of sources be negative [29]. Accordingly, if the sks is positive, the choice of negative cubic non-linearities, in principle, guarantees the convergence. Table 7.1 displays the results, in the form "bias $\pm$ standard deviation", obtained for the angle estimated by those three methods for several source-pair distributions. The true angle is $\theta = 30^{\circ}$ and, inspired by EASI, equal values are chosen for the second and fourth-order adaption coefficients $\lambda = \mu = 10^{-3}$. The results were calculated from the last $5 \cdot 10^4$ out of a total of $5.5 \cdot 10^4$ performed iterations, that is, right after convergence. The first two rows correspond to source pairs with negative kurtosis sum (uniform–uniform: -2.4, uniform–Gaussian: -1.2), whereas in the last two the source kurtosis sum is positive (Laplacian–exponential: 9, exponential–Rayleigh: ≈ 6.25). The same mixture realizations are fed into all three methods. Note that EASI does not converge for the last two source pairs (which probably could have been prevented with the normalized version of the algorithm [29], although for the sake of a fairer comparison only non-normalized implementations of the methods were considered). For the standard deviation it is found that $\sigma_{\hat{\theta}_{\mathrm{adEML}}} < \sigma_{\hat{\theta}_{\mathrm{AROT}}} < \sigma_{\hat{\theta}_{\mathrm{EASI}}}$. Table 7.2 shows the outcome of a similar experiment, but this time with $\lambda = \mu = 10^{-4}$ and taking the last $5 \cdot 10^4$ estimates from a total of 10^5 iterations. The above relationship among the respective standard deviations seems to hold again, concluding that adEML's convergence is more precise than the two other methods'.

Regarding the convergence speed, it was observed that for the same adaption coefficient adEML converges considerably faster than EASI and AROT. This is directly related to the robust high convergence speed of the adEML estimator, as explained in Section 7.3.3.

METHOD → / SOURCE DISTRIBUTION ↓	adEML	EASI	AROT
Uniform–uniform	0.054 ± 0.591	0.033 ± 1.002	0.156 ± 0.762
Uniform–Gaussian	-0.180 ± 3.037	-0.100 ± 3.433	-0.045 ± 3.324
Laplacian–exponential	-0.138 ± 1.575	No Convergence	-0.540 ± 1.690
Exponential–Rayleigh	-0.267 ± 2.722	No Convergence	-0.376 ± 3.068

Table 7.1: Estimated-angle bias $\pm$ standard deviation for adaption coefficients $\lambda = \mu = 10^{-3}$, evaluated from the last $5 \cdot 10^4$ out of $5.5 \cdot 10^4$ iterations. All values are expressed in degrees.

METHOD → / SOURCE DISTRIBUTION ↓	adEML	EASI	AROT
Uniform–uniform	0.159 ± 0.151	0.157 ± 0.364	0.189 ± 0.179
Uniform–Gaussian	-0.819 ± 0.766	-0.880 ± 0.909	-0.842 ± 0.573
Laplacian–exponential	0.466 ± 0.357	0.457 ± 1.887	0.574 ± 0.411
Exponential–Rayleigh	0.634 ± 0.779	0.090 ± 2.715	0.759 ± 0.909

Table 7.2: Estimated-angle bias $\pm$ standard deviation for adaptation coefficients $\lambda = \mu = 10^{-4}$, evaluated from the last $5 \cdot 10^4$ out of 10^5 iterations. All values are expressed in degrees.

7.8.3 Tracking non-stationarities

Single estimator

A time-varying rotation angle is now applied to two uniformly distributed sources. Specifically, the true angle used is mathematically described by:

$$\theta_k\,(^{\circ}) = \begin{cases} -30 + 60\sin(2\pi k/4000), & 0 < k \leqslant 1000, \\ -30 + 30\,(k-1000)/1000, & 1000 < k \leqslant 2000, \\ 40 - 60\,(k-2000)/1000, & 2000 < k \leqslant 3000, \\ -20, & k > 3000, \end{cases} \tag{7.59}$$

so that θ_k exhibits two abrupt changes, one at $k = 1000$ and another one at $k = 2000$. This function is represented by the steady solid line of Figure 7.11. The oscillating solid line in that plot shows the angle estimated by the adEML from the set of resultant whitened observations, for $\mu = 2 \cdot 10^{-2}$, averaged over 100 independent source-realizations. For the sake of comparison, the angle obtained by the EASI method of [29] (without the pre-whitening part in its serial updating rule) with $\lambda = 5.5 \cdot 10^{-3}$ and over

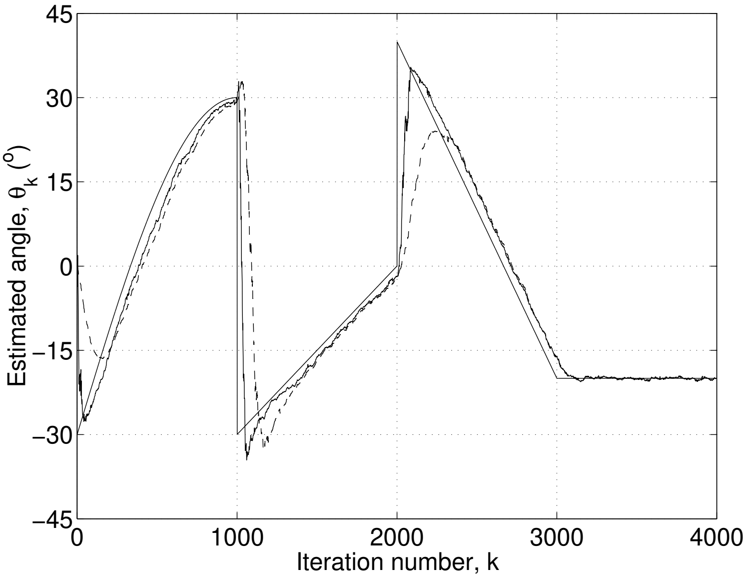


Figure 7.11: Non-stationary rotation angle. Two uniform sources. Steady solid line: true angle θ. Oscillating solid line: angle estimated by adEML, $\mu = 2 \cdot 10^{-2}$. Dashed line: angle estimated by EASI, $\lambda = 5.5{\cdot}10^{-3}$. Results are averaged over 100 independent MC realizations.

the same MC runs, is displayed in the dashed line of the same figure. The adaption coefficients of both methods were selected so that they produced the same variance of the estimated angle in steady state, in a bid to compare their convergence speed under the same accuracy conditions. It can be observed in Fig. 7.11 that adEML is up to more than twice as fast as EASI in tracking the time-varying parameter and recovering from its abrupt oscillations. Nevertheless, the adEML high-speed convergence characteristic noticed in Section 7.3.3 could be further exploited by means of a suitable abrupt-change detection scheme in order to yield a faster more efficient performance in the presence of such sudden variations. Essentially, the centroid should be taken to the origin of the complex plane as soon as the abrupt change is detected. Some abrupt-change detectors are studied in [11].

Combined estimation

We now assess the effectiveness of adaptive combined strategy adcombEML, characterized by decision rule (7.58), in a non-stationary environment. A constant unitary

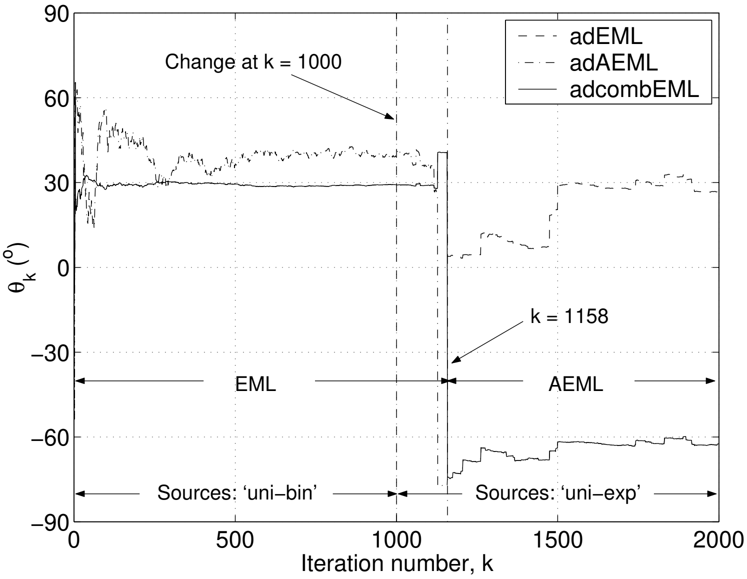


Figure 7.12: Illustration of the adcombEML method in non-stationary conditions, $\theta = 30^{\mathrm{o}}$, $\mu = 10^{-3}$. The plot shows the angle estimated by the adEML (dashed line), the adAEML (dash-dot line) and the adcombEML (solid line) methods versus time. Before $k = 1000$, the sources are composed of a uniform-binary pair. After $k = 1000$, the binary source is changed for an exponential signal. The switch from the adEML to the adAEML, governed by decision rule (7.58), is made at $k = 1158$.

transformation of $\theta = 30^{\mathrm{o}}$ is applied over two sources with uniform and binary distribution, respectively. The stationary conditions are broken at time instant $k = 1000$, when the binary source is changed for an exponential one. A constant adaption coefficient is chosen, $\mu = 10^{-3}$. Figure 7.12 shows the angle trajectories estimated by the adcombEML. The adEML is selected until $k = 1158$, instant at which the change is detected and the adAEML chosen, in compliance with decision rule (5.42)/(7.58). The standard deviations of the angle estimates before the switch are $\sigma_{\mathrm{adEML}} = 0.226^{\mathrm{o}}$ and $\sigma_{\mathrm{adAEML}} = 1.165^{\mathrm{o}}$ (evaluated between samples $k = 500$ and $k = 1000$), whereas after the change these values become $\sigma_{\mathrm{adEML}} = 1.751^{\mathrm{o}}$ and $\sigma_{\mathrm{adAEML}} = 0.837^{\mathrm{o}}$ (estimated between $k = 1500$ and $k = 2000$). Hence, switching proves beneficial, for the best performance is maintained, just as intended.

7.8.4 Performance of the full separation system

We illustrate the performance of the whole separation system by means of some simulations involving the adEML estimator. Analogous results are obtained for the other adaptive methods studied in this chapter.

As explained in Section 7.2, angle estimator (7.12) can be turned into a complete separation system just by combining it with an adaptive pre-whitening stage. In the following simulations, the serial updating rule (7.6) (EASI's original pre-whitening rule) is used to obtain adaptively the whitening matrix and generate the corresponding samples of the whitened observations, from which the adEML estimator can then be employed in order to find the missing rotation. The same pre-whitening strategy is used for AROT. The incorporation of pre-whitening allows us, in turn, to pay attention to the impact of the mixing-matrix conditioning on the algorithms' convergence, along the lines of [56].

Two signals

A two-sensor mixture is made up of a uniformly-distributed source, a binary source and a regular mixing matrix with condition number $n_{\mathrm{M}} = 10$.[9] We choose the step sizes $\lambda = \mu = 10^{-2}$. The trajectories of the modulus of the elements of the global mixing-unmixing system (which relates true and estimated sources; see Section 7.2 and Fig. 7.1) by the adEML averaged over 100 independent source-realizations are plotted in the solid lines of Fig. 7.13. Two of the elements converge to 1, whereas the other two do so to 0, thus achieving a successful source separation. The same mixture realizations are processed by EASI [this time including the pre-whitening stage (7.6) too] with positive cubic non-linearities and identical adaption coefficient. The selected non-linearities fulfil the stability condition of this method for the mentioned kind of sources. The averaged global matrix elements obtained by EASI are represented in Fig. 7.14 (solid lines). The results obtained by AROT under exactly the same conditions appear in Fig. 7.15 (solid lines). All three procedures converge nearly as fast, with minor differences in their oscillations once convergence is reached (at about the 500th iteration). adEML and EASI results are quite similar, whereas AROT seems to exhibit more difficulties in the

[9] The condition number n_{M} is defined as the ratio of the largest to the smallest singular value of M.

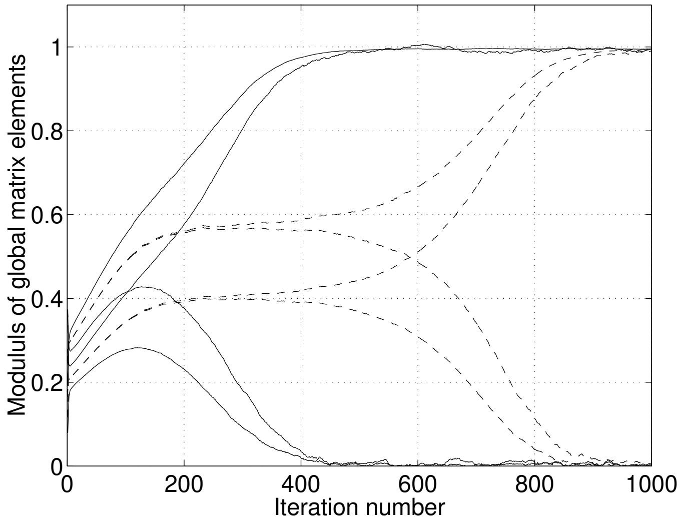


Figure 7.13: Evolution of the global system averaged over 100 independent realizations in the separation of a uniform and a binary source from two sensors by the adEML method, $\lambda = \mu = 10^{-2}$, with mixing-matrix condition number n_{M}. Solid lines: $n_{\mathrm{M}} = 10$. Dashed lines: $n_{\mathrm{M}} = 10^3$.

transient period, until the moments are estimated with enough accuracy. The inclusion of the pre-whitening stage accounts for the decrease in speed of the adEML method, relative to the results of the preceding sections. A theoretic study of the influence of pre-whitening on the method's performance is required to confirm this hypothesis. However, space is lacked here for such thorough analysis.

On the other hand, the impact of the mixing-matrix conditioning is evidenced by the dashed lines of Figs. 7.13, 7.14 and 7.15, obtained for $n_{\mathrm{M}} = 10^3$. Convergence is delayed as the mixing-matrix conditioning worsens.

More than two signals

In this scenario, the extension described in Section 7.5 and Fig. 7.7 is used for the adEML method. On the part of the AROT method, only a single sweep is performed over the signal pairs, following the guidelines of [41]. A fixed regular (3×3) mixing matrix is chosen, with $n_{\mathrm{M}} = 10$, which generates a set of three mixtures when applied on a set of three independent sources: a binary sequence, a uniformly distributed signal and a sinusoid with random frequency and initial phase. The adaption coefficients

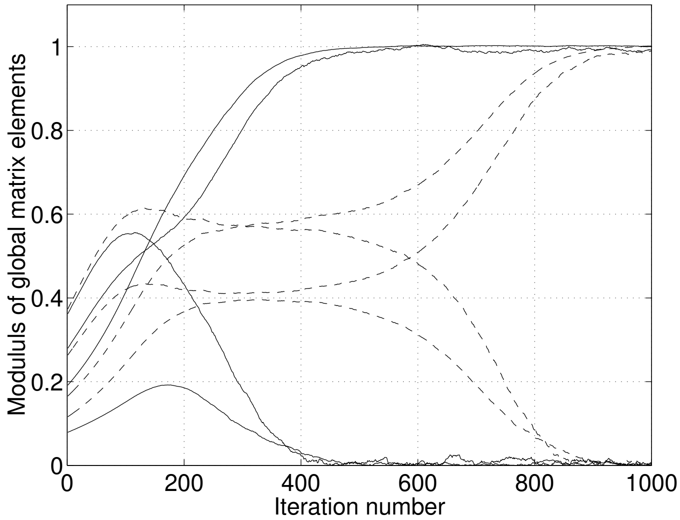


Figure 7.14: Global system trajectories averaged over 100 independent realizations in the separation of a uniform and a binary source from two observations by the EASI method, cubic non-linearities, $\lambda = 10^{-2}$, with mixing-matrix condition number n_{M}. Solid lines: $n_{\mathrm{M}} = 10$. Dashed lines: $n_{\mathrm{M}} = 10^3$.

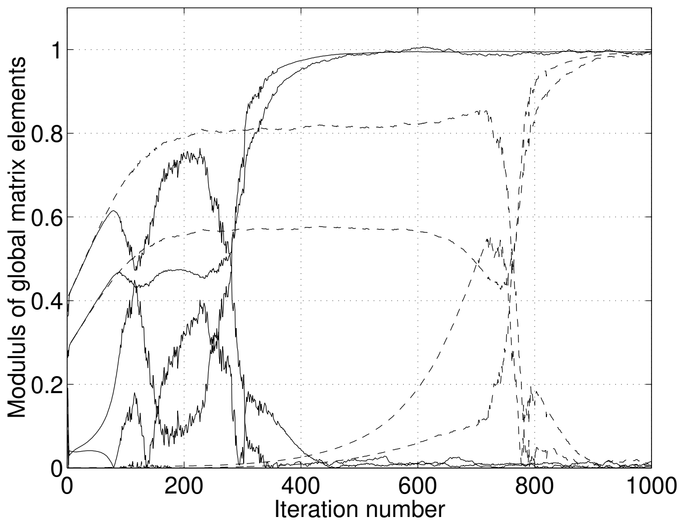


Figure 7.15: Global matrix coefficients averaged over 100 independent realizations in the separation of a uniform and a binary source from two measurements by the AROT method, $\lambda = \mu = 10^{-2}$, with mixing-matrix condition number n_{M}. Solid lines: $n_{\mathrm{M}} = 10$. Dashed lines: $n_{\mathrm{M}} = 10^3$.

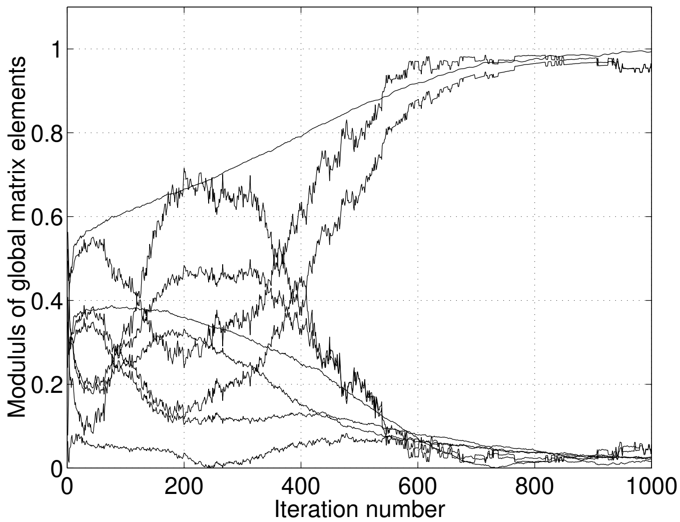


Figure 7.16: Modulus of global system coefficients averaged over 100 mixture realizations. Three-source (binary–uniform–sinusoid) three-sensor scenario. adEML method, $\lambda = \mu = 4.5 \cdot 10^{-3}$, mixing-matrix condition number $n_{\mathrm{M}} = 10$.

are now selected as $\lambda = \mu = 4.5 \cdot 10^{-3}$ for all three methods. Figs. 7.16 and 7.17 display the trajectories of the global system entries obtained by adEML and EASI, respectively, averaged over 100 independent MC realizations. AROT's results are not shown since the procedure did not converge on 20% of the MC runs (perhaps this negative outcome could have been prevented with some sort of normalization in the algorithm; this could be a subject of further investigations). Looking at Fig. 7.16 one can see that the proposed extension works fine, a conclusion which is supported by additional experiments. EASI's results are smoother than adEML's, but the former method consistently converged to a non-separating solution, in which one of the sources was extracted but the other two remained mixed at the separator output. The influence of the mixing-matrix conditioning on the convergence was observed to be analogous to that in the two-signal case.

7.9 Summary and Conclusions

This chapter has looked into adaptive BSS. After reviewing the basic aspects revolving around the adaptive problem, our interest has focused on the adaptive implementations

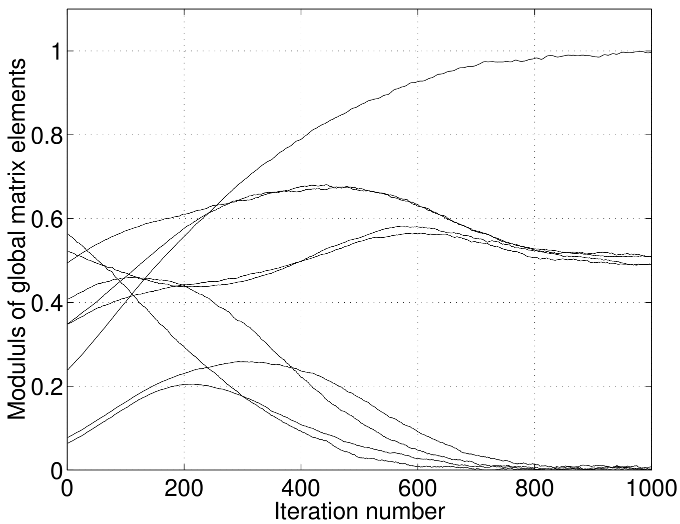


Figure 7.17: Modulus of global system coefficients averaged over 100 mixture realizations. Three-source (binary–uniform–sinusoid) three-sensor scenario. EASI method, cubic non-linearities, $\lambda = 4.5 \cdot 10^{-3}$, mixing-matrix condition number $n_{\mathrm{M}} = 10$.

of the EML, AEML and the combEML estimators. For a variety of reasons (specially, robust, extremely fast convergence and good accuracy), the implementations based on the direct adaptive updating of the centroids has been favoured over the procedures derived from the optimization of the respective cost functions. The simple algebraic formalism on which these adaptive methods are based has allowed us to capitalize on the standard devices for the analysis of stochastic algorithms when studying the properties of the methods. A detailed analysis of their convergence and properties has showed that the adEML and the adAEML present (in the i.i.d. case) global stable convergence to a separation solution under the same applicability conditions of their batch counterparts, that is, if the sks and the skd, respectively, are not null. This attribute is in sharp contrast with other adaptive methods (e.g., [29,75]), which are not guaranteed to converge to a desired solution. The analysis has culminated in the derivation of their respective asymptotic pdf's, together with their Gaussian approximations, which allow the assessment of the estimation quality once convergence is reached. Simulations have shown the accuracy of these asymptotic results, and have revealed the estimators' high speed in tracking non-stationarities of the mixing system. By using a pairwise extension of computational cost comparable to other adaptive procedures, the methods have been made applicable in BSS contexts of more than two signals, which has been empirically supported. Experiments have also demonstrated the use of the estimators in a complete separation scheme (including pre-whitening as well), revealing a performance up to the mark of other adaptive BSS methods.

Chapter 8

Extensions

8.1 Introduction

Up to now, only estimators based on the 4th-order cumulants of the data have been considered. This is not an isolated situation. On assorted grounds, most references in the literature are concerned with methods based on 4th-order statistics. Third-order cumulants are often disregarded on the account of most practical source distributions being symmetric, hence rendering cumulants of such order ill-conditioned [42]. Orders higher than four are also overlooked because the estimation errors as well as the computational requirements generally increase as the order rises. Order two, as we have seen, lacks the ability to achieve the separation when i.i.d. signals are processed, or any time coherence is just ignored. Order four, therefore, emerges as the most reasonable compromise. Nevertheless, perhaps spurred by an indomitable scientific curiosity, we dare explore in Section 8.2 estimators based on the nth-order statistics of the whitened observations, for n arbitrary. It is anticipated that a number of appealing, unexpected results will be obtained, which, in some sense, can be considered as an extension to orders other than four of the results presented in the previous chapters.

On the other hand, in certain applications such as digital communications [52] and seismic sounding [83], signals are more conveniently treated by means of complex-valued representations (phasors). As a consequence, the BSS problems arising in such environments correspond to complex-valued mixtures of complex-valued source signals. Aiming to tackle the complex-valued problem, the extension of some of the ideas developed in this thesis to the complex-signal case is accomplished in Section 8.3.

8.2 Arbitrary-Order Cumulants

In this section, the EML estimator (3.32) is found to be a particular case, at order 4, of a wider family of closed-form estimators based on the nth-order cumulants of the whitened output. Section 8.2.1 presents this novel family of analytic expressions, while Section 8.2.2 develops their respective optimization criteria. At $n = 3$, a novel third-order estimator is deduced from such family. The analysis of this new estimator is accomplished in Section 8.2.3. Simulations reported in Section 8.2.4 illustrate its behaviour.

8.2.1 A family of closed-form estimators

We have seen that the EML centroid can be written as a particular complex linear combination of the whitened-vector 4th-order cumulants [eqn. (3.34)]. When developing such combination, it turns out that the centroid modulus corresponds to another combination of the source 4th-order cumulants (the absolute sks), whereas the centroid phase is directly linked to the unknown parameter [eqn. (3.36)]. Inspired by this observation, we now induce the specific complex linear combinations of the whitened-output cumulants, at any order, which are able to preserve, like the EML, explicit expressions for the parameter of interest. The following result provides such extension.

Theorem 8.1. *Define $\xi_n(\mathbf{z})$ as the following complex weighted sum of pairwise nth-order cumulants of the components of $\mathbf{z}$, for $n \in \mathbb{N}^+$:*

$$\xi_n(\mathbf{z}) \triangleq \sum_{r=0}^{n} \binom{n}{r} j^r \kappa^z_{n-r,r}. \tag{8.1}$$

If $\mathbf{z} = \mathcal{Q}(\theta)\,\mathbf{x}$, with $\mathbf{x}$ made up of independent components, then

$$\xi_n(\mathbf{z}) = \mathrm{e}^{jn\theta}\xi_n(\mathbf{x}), \tag{8.2}$$

where, according to (8.1),

$$\xi_n(\mathbf{x}) = \kappa^x_{n0} + j^n \kappa^x_{0n}. \tag{8.3}$$

Proof. Due to the multilinearity property of cumulants and the statistical independence of the components in $\mathbf{x}$ (see Appendices A.2– A.3), we can write:

$$\kappa^z_{n-r,r} = \sum_{i_1 \ldots i_n} q_{1i_1} \ldots q_{1i_{n-r}} q_{2i_{n-r+1}} \ldots q_{2i_n} \mathrm{Cum}^x_{i_1 \ldots i_n} = q_{11}^{n-r} q_{21}^r \kappa^x_{n0} + q_{12}^{n-r} q_{22}^r \kappa^x_{0n}. \tag{8.4}$$

Substituting this expression into (8.1) and noting that $q_{11} = \cos\theta$, $q_{12} = -\sin\theta$, $q_{21} = \sin\theta$ and $q_{22} = \cos\theta$, we obtain:

$$\begin{aligned} \xi_n(\mathbf{z}) &= \sum_{r=0}^{n} \binom{n}{r} j^r \left[(\cos\theta)^{n-r} (\sin\theta)^r \kappa^x_{n0} + (-\sin\theta)^{n-r} (\cos\theta)^r \kappa^x_{04} \right] = \\ &= \kappa^x_{n0} \sum_{r=0}^{n} \binom{n}{r} (\cos\theta)^{n-r} (j\sin\theta)^r + j^n \kappa^x_{0n} \sum_{r=0}^{n} \binom{n}{r} (j\sin\theta)^{n-r} (\cos\theta)^r, \end{aligned} \tag{8.5}$$

where, to go from the second to the third term, we have made use of the relations $j^r = j^n j^{-(n-r)} = j^n (j^{-1})^{n-r} = j^n (-j)^{n-r}$. Now, in order to simplify the summations we resort to Newton's binomial expansion and Euler's formula, yielding:

$$\sum_{r=0}^{n} \binom{n}{r} (\cos\theta)^r (j\sin\theta)^{n-r} = \sum_{r=0}^{n} \binom{n}{r} (\cos\theta)^{n-r} (j\sin\theta)^r = (\cos\theta + j\sin\theta)^n = \mathrm{e}^{jn\theta}. \tag{8.6}$$

So, finally:

$$\xi_n(\mathbf{z}) = \mathrm{e}^{jn\theta} (\kappa^x_{n0} + j^n \kappa^x_{0n}) = \mathrm{e}^{jn\theta} \xi_n(\mathbf{x}). \tag{8.7}$$

The last equality is due to the source independence, which makes all source cross-cumulants vanish. ■

The significance of the result stated by Theorem 8.1 lies in the fact that, from (8.2), the hidden rotation can be estimated in analytic form as:

$$\hat{\theta}_n = \frac{1}{n} \arg\left(\frac{\xi_n(\mathbf{z})}{\xi_n(\mathbf{x})} \right). \tag{8.8}$$

If the source statistics at order n are known, equation (8.8) provides a closed-form estimator of the sought parameter θ. However, even if the source statistics are unknown, which is the typical case in a genuine blind problem, the analytic expression still proves useful. This is because in some situations $\xi_n(\mathbf{x})$ may be estimated from the available data, as illustrated later. Formula (8.8), nevertheless, must be applied with some care.

Solution indeterminacy

As $n\,\theta = n(\theta + 2\pi m/n)$, for any $n \in \mathbb{N}^+$, $m \in \mathbb{Z}$, it turns out that the actual angle value supplied by estimator (8.8) at order n can be any of the form $\hat{\theta}_n = \theta + 2\pi m/n$, $m \in \mathbb{Z}$. The only order at which this solution indeterminacy is acceptable is, in principle, $n = 4$, since then the possible bias takes the form $\pm m\pi/2$, and this still recovers the source waveforms up to order permutation and unitary scale factors (allowable indeterminacies, as commented in Sections 2.1.4 and 3.2.1). Accordingly, the indeterminacy is also admissible at second order, but we will see next that at $n = 2$ eqn. (8.8) is not actually of any use. At order $n = 3$, it will also be seen later than the indeterminacy can be resolved. At any order $n > 4$, there is no other option than considering together the set of possible solutions $\hat{\theta}_n + 2\pi m/n$, $m = 0, 1, \ldots, n-1$, and deciding by means of an extra criterion (such as degree of output statistical independence) which of those n values is correct. In addition, remark that eqn. (8.8) only provides estimates in the range $\Omega_{2\pi/n} \triangleq [-\pi/n, \pi/n]$. Therefore, from a practical point of view, estimators (8.8) are only useful as actual direct methods for $n = 3$ and $n = 4$.

Theorem 8.1 will be extended to the complex case in Section 8.3. Estimation scheme (8.8) is now studied at different orders.

Order 2

At order 2 we have:

$$\xi_2(\mathbf{z}) = \mathrm{e}^{j2\theta}(\kappa_{20}^x - \kappa_{02}^x) = 0, \tag{8.9}$$

owing to the source unit-variance convention. Thus, no meaningful estimation can be carried out by (8.9) at this order. This was to be expected, since all second-order information is depleted after pre-whitening. Indeed, no second-order based technique is able to recover the missing orthogonal transformation if the temporal structure cannot be exploited (as in the i.i.d. case) or is just ignored [138] (see also Section 2.2). At orders 3 and 4, the situation is quite different. We will see that in those cases estimator (8.8) can be used without the need for any prior knowledge of the source statistics.

Order 3

In this case:

$$\xi_3(\mathbf{z}) = \mathrm{e}^{j3\theta}(\kappa_{30}^x - j\kappa_{03}^x). \tag{8.10}$$

In principle, we would need to know κ_{30}^x and κ_{03}^x in order to estimate θ through (8.10) [in the following, for simplicity we just write ξ_3 for $\xi_3(\mathbf{z})$]:

$$\hat{\theta}_3 = \frac{1}{3}\arg\left(\frac{\xi_3}{\gamma_3}\right), \qquad \gamma_3 \triangleq \kappa_{30}^x - j\kappa_{03}^x, \tag{8.11}$$

but then this estimator would still be suffering from the solution indeterminacy commented above. Nevertheless, a closer look at the 3rd-order cumulants of $\mathbf{z}$ reveals (resorting again to the multilinearity property) that:

$$\kappa_{21}^z + j\kappa_{12}^z = \tfrac{1}{2}\sin 2\theta\, \mathrm{e}^{j\theta}\gamma_3 = \gamma_3'. \tag{8.12}$$

Hence, substituting (8.12) into (8.10): $\xi_3/\gamma_3' = \frac{2}{\sin 2\theta}\,\mathrm{e}^{j2\theta}$. As a result, in practice θ may be estimated at order 3 via:

$$\hat{\theta}_3 = \frac{1}{2}\arg\left(\frac{\hat{\xi}_3}{\hat{\gamma}_3'}\right), \tag{8.13}$$

in which $\hat{\xi}_3$ and $\hat{\gamma}_3'$ are the respective sample estimates of their population counterparts, and choosing $\hat{\theta}_3 = 0$ if $\hat{\gamma}_3' = 0$. This is called *Third-Order Blind Signal Estimator (TOBSE)*. Note that the possible negative sign introduced by the term $\sin 2\theta$ of γ_3' is irrelevant, as it would result in a $\pm\pi/2$ radian bias in θ which does not affect the source recovery. Also, the solution indeterminacy is avoided, since the argument of ξ_3/γ_3' depends on twice the true angle θ. Naturally, this 3rd-order estimator is applicable as long as there is at least one asymmetrically distributed source. If this condition holds, then it may be wiser to employ this 3rd-order estimator rather than a 4th-order one, since given the same number of samples the 3rd-order cumulant estimates would always offer higher accuracy than the 4th-order cumulant estimates. This notion will be endorsed by the experimental results in Section 8.2.4. Finally, remark that the computation of the 3rd-order statistics ξ_3 and γ_3' can be done in polar form, similarly as for the AML (Section 3.3.1), EML (Section 3.3.2) and AEML (Section 5.2.1) estimators:

$$\xi_3 = \mathrm{E}[(z_1 + jz_2)^3] = \mathrm{E}[r^3\mathrm{e}^{j3\phi}] \tag{8.14}$$

$$\gamma_3' = \mathrm{E}[z_1 z_2(z_1 + jz_2)] = \mathrm{E}[\tfrac{1}{2}\, r^3\mathrm{e}^{j\phi}\sin 2\phi]. \tag{8.15}$$

The TOBSE is analyzed in Section 8.2.3.

We will call estimator (8.11) *TONBSE*, where "NB" stands for non-blind, as it requires the source skewness to be known in advance.

Order 4

For $n = 4$, expression (8.2) develops into:

$$\xi_4(\mathbf{z}) = (\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z) + j4(\kappa_{31}^z - \kappa_{13}^z) = \mathrm{e}^{j4\theta}(\kappa_{40}^x + \kappa_{04}^x) = \mathrm{e}^{j4\theta}\gamma. \tag{8.16}$$

This is precisely the EML centroid ξ (3.21). Again, we do not need to have any prior knowledge of the source distribution, since the sks γ may be estimated from the whitened signals as in eqns. (3.26) and (3.27). As $\gamma \in \mathbb{R}$, only its sign is of interest for the estimation of θ.

8.2.2 Cumulant-matching and cost-function optimization criteria

The family of estimators defined by equation (8.8) can be seen as resulting from the optimization of respective cumulant matching and cost function criteria. Before unveiling such criteria, we first need the following corollary of Theorem 8.1, which is an extension to order n of Lemma 4.2.

Corollary 8.2. Define the output vector as a counter-rotation of angle $\hat{\theta}$ performed on the whitened signal space, i.e., $\mathbf{s} = \mathcal{Q}(-\hat{\theta})\,\mathbf{z}$. Then, with the relationships introduced in Theorem 8.1:

$$\xi_n(\mathbf{s}) = \mathrm{e}^{-jn\hat{\theta}}\xi_n(\mathbf{z}). \tag{8.17}$$

In particular, a rotation does not alter the nth-order centroid modulus: $|\xi_n(\mathbf{s})| = |\xi_n(\mathbf{z})| = |\xi_n(\mathbf{x})|$.

Proof. $\mathbf{s} = \mathcal{Q}(-\hat{\theta})\,\mathbf{z} = \mathcal{Q}(-\hat{\theta})\mathcal{Q}(\theta)\,\mathbf{x} = \mathcal{Q}(\theta - \hat{\theta})\,\mathbf{x}$, so, by successive application of Theorem 8.1: $\xi_n(\mathbf{s}) = \mathrm{e}^{jn(\theta-\hat{\theta})}\xi_n(\mathbf{x}) = \mathrm{e}^{-jn\hat{\theta}}\big(\mathrm{e}^{jn\theta}\xi_n(\mathbf{x})\big) = \mathrm{e}^{-jn\hat{\theta}}\xi_n(\mathbf{z})$. ■

Theorem 8.3. *The family of estimators given analytically by equation* (8.8) *are the local minimizers of*

$$J_{\mathrm{m}n}(\hat{\theta}) = \big|\xi_n(\mathbf{s}) - \xi_n(\mathbf{x})\big|^2 = \big|\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}} - \xi_n(\mathbf{x})\big|^2, \tag{8.18}$$

or, equivalently, the local maximizers of

$$J_{\mathrm{M}n}(\hat{\theta}) = \mathrm{Re}\big(\xi_n(\mathbf{x})^*\xi_n(\mathbf{s})\big) = \mathrm{Re}\big(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}}\big), \tag{8.19}$$

that is,

$$\hat{\theta}_n = \arg\min_{\hat{\theta}} J_{\mathrm{m}n} = \arg\max_{\hat{\theta}} J_{\mathrm{M}n}. \tag{8.20}$$

Proof. From Corollary 8.2: $J_{\mathrm{m}n}(\hat{\theta}) = \big(\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}} - \xi_n(\mathbf{x})\big)\big(\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}} - \xi_n(\mathbf{x})\big)^* = |\xi_n(\mathbf{z})|^2 + |\xi_n(\mathbf{x})|^2 - 2\mathrm{Re}\big(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}}\big) = 2|\xi_n(\mathbf{x})|^2 - 2J_{\mathrm{M}n}(\hat{\theta})$. In addition, $|\xi_n(\mathbf{x})|^2$ is constant for a given pair of sources, and hence the minimization of $J_{\mathrm{m}n}$ is tantamount to the maximization of $J_{\mathrm{M}n}$. In consequence, only the second part of the theorem needs to be proven.

The first and second derivatives of $J_{\mathrm{M}n}$ are, respectively:

$$\frac{\partial J_{\mathrm{M}n}}{\partial \hat{\theta}} = n\,\mathrm{Im}\big(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}}\big), \qquad \frac{\partial^2 J_{\mathrm{M}n}}{\partial \hat{\theta}^2} = -n^2\mathrm{Re}\big(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z})\,\mathrm{e}^{-jn\hat{\theta}}\big) = -n^2 J_{\mathrm{M}n}. \tag{8.21}$$

The local extrema of $J_{\mathrm{M}n}$ are given by the first-order necessary condition:

$$\frac{\partial J_{\mathrm{M}n}}{\partial \hat{\theta}} = 0 \;\Leftrightarrow\; \mathrm{e}^{j[\angle(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z}))-n\hat{\theta}]} = \pm 1. \tag{8.22}$$

The local maxima must also fulfil the second-order necessary condition $\frac{\partial J_{\mathrm{M}n}}{\partial \hat{\theta}} < 0$, which occurs if and only if $\mathrm{e}^{j[\angle(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z}))-n\hat{\theta}]} = 1$. Hence $\hat{\theta}_{\mathrm{opt}} = \frac{1}{n}\angle\big(\xi_n(\mathbf{x})^*\xi_n(\mathbf{z})\big) + \frac{2\pi m}{n}$, $m \in \mathbb{Z}$, which transforms into $\hat{\theta}_n$ with $1/\xi_n(\mathbf{x}) = \xi_n(\mathbf{x})^*/|\xi_n(\mathbf{x})|^2$. ■

Minimization of (8.18) constitutes a cumulant matching criterion, whereby the mismatch between a linear combination of output nth-order cumulants $(\xi_n(\mathbf{s}))$ and the same linear combination of source cumulants $(\xi_n(\mathbf{x}))$ is minimized. Maximization of (8.19) yields an equivalent cost function optimization criterion. In the light of Theorem 8.3, the optimization of these equivalent criteria is solved in closed-form by (8.8). Note that at $n = 4$, all these extremization principles reduce to the criteria found for the EML estimator in Chapter 4.

8.2.3 Analysis of the third-order estimator

In the following sections we study the asymptotic performance, the computational complexity and the extension to more than two signals of the TOBSE estimator.

Asymptotic pdf and variance

Due to the term $\hat{\gamma}_3'$ in the denominator of (8.13), carrying out the analysis of the TOBSE estimator is a difficult task. In a bid to simplify this study, we assume that γ_3 is known (or perfectly estimated), and analyze the TONBSE estimator (8.11) directly. In this manner, a higher (i.e., optimistic or 'best-case') bound of performance is obtained.

The sample estimate of (8.14) from a block of T whitened samples is given by:

$$\hat{\xi}_3 = \frac{1}{T}\sum_{k=1}^{T}\big(z_1(k)+jz_2(k)\big)^3 = \left\{\frac{1}{T}\sum_{k=1}^{T}\big(x_1(k)+jx_2(k)\big)^3\right\}\mathrm{e}^{j3\theta} = \hat{\xi}_{3_x}\mathrm{e}^{j3\theta}, \tag{8.23}$$

with $\hat{\xi}_{3_x} \triangleq \frac{1}{T}\sum_{k=1}^{T}\big(x_1(k)+jx_2(k)\big)^3$, and where the second equality holds by virtue of (3.14). Accordingly, the estimation error can be expressed as:

$$\Delta\hat{\theta}_3 \triangleq \hat{\theta}_3 - \theta = \frac{1}{3}\angle(\hat{\xi}_{3_x}\gamma_3^*). \tag{8.24}$$

Now, the expansion of $\hat{\xi}_{3_x}\gamma_3^*$ yields:

$$\begin{aligned}\hat{\xi}_{3_x}\gamma_3^* = \frac{1}{T}\sum_k\Big\{&\big[\kappa_{30}^x(x_1^3-3x_1x_2^2)-\kappa_{03}^x(3x_1^2x_2-x_2^3)\big]\\ &+j\big[\kappa_{30}^x(3x_1^2x_2-x_2^3)+\kappa_{03}^x(x_1^3-3x_1x_2^2)\big]\Big\} = \frac{1}{T}\sum_k(\tilde{\omega}_1+j\tilde{\omega}_2).\end{aligned} \tag{8.25}$$

Thus, we have been able to write the above expression in the terms used in the asymptotic analyses of Sections 3.3.5, 3.4.3 and 5.2.3. Therefore, the asymptotic pdf has the same shape as (3.66), now with $\delta = \angle(\hat{\xi}_{3_x}\gamma_3^*) = 3\Delta\hat{\theta}_3$. The corresponding pdf parameters are in this case:

$$\mathrm{E}[\tilde{\omega}_1] = (\kappa_{30}^x)^2 + (\kappa_{03}^x)^2 = |\gamma_3|^2 \tag{8.26}$$

$$\mathrm{E}[\tilde{\omega}_2] = 0 \tag{8.27}$$

$$\mathrm{E}[\tilde{\omega}_1^2] = (\kappa_{30}^x)^2(\mu_{60}^x+9\mu_{04}^x-6\mu_{40}^x) + (\kappa_{03}^x)^2(\mu_{06}^x+9\mu_{40}^x-6\mu_{04}^x) + 20(\kappa_{30}^x\kappa_{03}^x)^2 \tag{8.28}$$

$$\begin{aligned}\mathrm{E}[\tilde{\omega}_2^2] = \tilde{\sigma}_2^2 = {}&(\kappa_{30}^x)^2(\mu_{06}^x+9\mu_{40}^x-6\mu_{04}^x) + (\kappa_{03}^x)^2(\mu_{60}^x+9\mu_{04}^x-6\mu_{40}^x)\\ &-20(\kappa_{30}^x\kappa_{03}^x)^2.\end{aligned} \tag{8.29}$$

So the asymptotic variance is then:

$$\mathrm{Var}[\hat{\theta}_3] \xrightarrow[T\to\infty]{} \sigma^2_{\text{TOBSE}} = \frac{\tilde{\sigma}_2^2}{9T|\gamma_3|^4}, \tag{8.30}$$

the term $\tilde{\sigma}_2^2$ being given by (8.29).

METHOD	EML	CF/ACF	AEML	TOBSE
Scatter-plot based form	$8T$	—	$6T$	$10T$
Cumulant based form	$8T$	$5T$	$7T$	$6T$

Table 8.1: Computational complexity of several closed-form BSS estimators per signal pair processed, with signals composed of T samples.

Small-error analysis of the cost function

At third order, cost function $J_{\mathrm{M}n}$ (8.19) becomes: $J_{\mathrm{M}3} = J_{\mathrm{TOBSE}} = \mathrm{Re}(\gamma_3^* \xi_3)$. The small-error analysis of this function (accomplished along the same lines of Sections 4.5 and 5.2.4) provides an asymptotic MSE equal to the variance obtained above. Effectively, calling $\hat{J}$ the sample equivalent of cost function J_{TOBSE}, we have:

$$\hat{J}'(\theta) \xrightarrow[T\to\infty]{\mathrm{d}} \mathrm{N}\left(0, \frac{9\tilde{\sigma}_2^2}{T}\right) \tag{8.31}$$

$$\hat{J}''(\theta) \xrightarrow[T\to\infty]{\mathrm{wp1}} -9|\gamma_3|^2, \tag{8.32}$$

leading to the mentioned result:

$$\mathrm{MSE}_3 \triangleq \mathrm{E}[(\hat{\theta}_3 - \theta)^2] \approx \frac{\tilde{\sigma}_2^2}{9T|\gamma_3|^4}, \tag{8.33}$$

for T sufficiently high.

Computational complexity

Following the guidelines of Sections 3.3.6, 3.4.4 and 5.2.5, the analysis of the computational requirements of the TOBSE estimator yields, for sample size T:

1. From the scatter-plot based expressions, eqns. (8.14) and (8.15): $10T$ flops.

2. From the cumulants, eqns. (8.10) and (8.12): $6T$ flops.

The cumulant form is notably less expensive that the more compact complex-centroid form. Table 8.1 condenses these results, together with the computational cost of the other closed-form BSS estimators studied in this thesis.

Extension to more than two signals

The pairwise extension described in Section 6.4.1 was not satisfactory for TOBSE. We believe that the reason for this negative outcome lies in the performance deterioration of the TOBSE estimation for true angle values around 0^o (see Section 8.2.4). Near convergence, the remaining true angles approach zero. As a consequence, these angles are not accurately estimated by TOBSE, and convergence is never reached. More investigations are needed to circumvent this difficulty, thus making the method applicable to the general BSS scenario.

8.2.4 Simulations

A few computer experiments assess the 3rd-order estimators derived and examined in the previous sections.

Variation with angular parameter

Figure 8.1 shows the performance variation of the TOBSE and TONBSE estimators as a function of the true parameter θ. The sources have exponential–uniform distributions, and are composed of $T = 5 \cdot 10^3$ samples. We appreciate that TOBSE degrades around the critical region $\theta \equiv 0$. This dependence, which is analogous to that observed for the CF/ACF estimator (Section 3.4.5), is probably introduced by the term $\hat{\gamma}'_3$. On the other hand, TONBSE is orthogonal invariant but, as anticipated in Section 8.2.1 (see 'Solution indeterminacy'), only provides meaningful estimation in the range $\Omega_{2\pi/3} = [-\pi/3, \pi/3]$.

Fitness of asymptotic variance/MSE

The accuracy of asymptotic variance/MSE approximations (8.30)/(8.33) is tested in Figure 8.2. Note how precisely the asymptotic values fit, even for as low a sample size as $T = 10$, the results by TOBSE and TONBSE when θ is outside their respective critical regions (defined by Fig. 8.1). For θ near the boundaries of their critical regions, the performance degrades, which is not predicted by the thoretical asymptotic results.

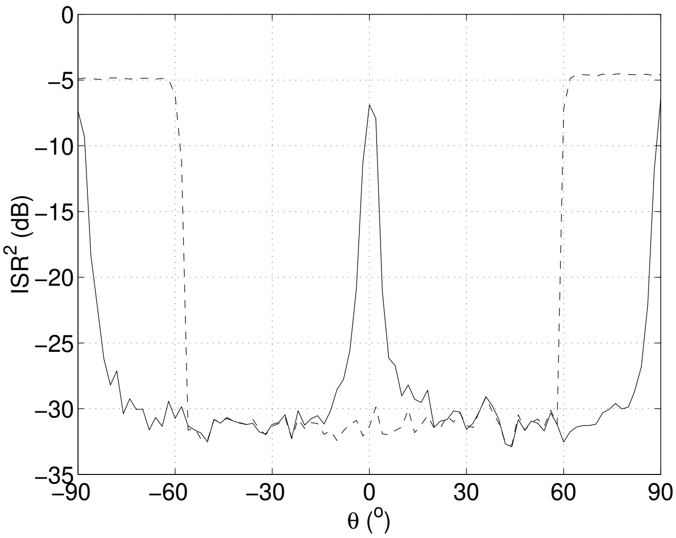


Figure 8.1: Performance of TOBSE (solid line) and TONBSE (dashed line) as a function of the true parameter θ, for exponential–uniform sources. $T = 5 \cdot 10^3$ samples, 100 MC runs.

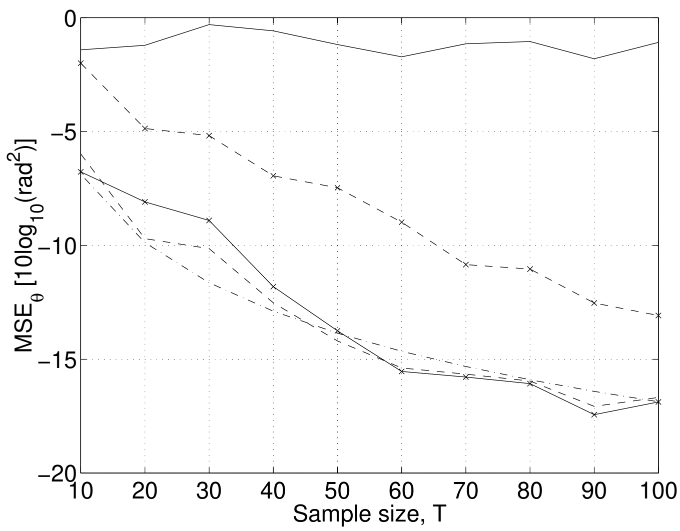


Figure 8.2: Performance of TOBSE (solid lines) and TONBSE (dashed lines) as a function of the sample size, for exponential–uniform sources. ν MC runs with $\nu T = 5 \cdot 10^5$. No markers: $\theta = 5^{\rm o}$; '$\times$'-markers: $\theta = 45^{\rm o}$. The dash-dot line represents asymptotic approximation (8.33).

Performance for several source distributions

TOBSE's performance for a number of source distribution combinations is assessed in Table 8.2, which shows the theoretic asymptotic values of $T\sigma^2_{\text{TOBSE}}$. The figures are calculated via eqn. (8.30), together with the population statistics of Tables A.1 and A.2. Bolface figures denote those source combinations for which TOBSE outperforms both EML and AEML (cf. Table 5.1). This occurs in 36.11% (13 out of 36) of all combinations, and in 61.90% (13 out of 21) of combinations with at least one asymmetric source. Observe that, as discussed above, the values in Table 8.2 would only be obtained for θ not close to 0.

Variation with source skewness and comparison with 4th-order estimators

A final simulation is intended to compare TOBSE with other estimators studied in the previous chapters, namely EML, AEML and combEML. We also aim to test how TOBSE's performance alters when the source skewness varies. These two objectives can be jointly achieved by resorting to the ideas of Appendix B. Two PRBS signals play the role of sources. One is kept at kurtosis $\kappa^x_{40} = -2$, and hence at skewness $\kappa^x_{30} = 0$. The kurtosis of the other source, κ^x_{04}, is gradually swept between -2 and 12, so that its skewness, κ^x_{03}, varies between 0 and -3.74. Figure 8.3 shows the results. TOBSE's performance appears very stable with the source skewness value. For θ near 0, a degradation in performance is observed around $\kappa^x_{03} = 0$, which was expected since in that case (both sources symmetric) the 3rd-order estimator is not applicable. But, surprisingly, this deterioration does not occur for θ far from such critical region. With PRBS sources TOBSE outperforms EML for $\gamma > -2$, and AEML for $\gamma < -2$, but is outperformed by combEML practically throughout the whole sks range.

8.3 Complex Mixtures

Our next purpose is to extend some of the results previously unfolded to the complex-mixture case. In particular, we want to lay emphasis on the existing analogy between the direct solutions to the real and the complex problem, which has not been evidenced to date. The definition of a special set of numbers in accordance with the particular

		$\kappa_{40}^x < 0$				$\kappa_{40}^x = 0$	$\kappa_{40}^x > 0$		
		Bin	Sin	Uni	Tri	Gau	Ray	Lap	Exp
$\kappa_{04}^x < 0$	Bin	∞	∞	∞	6.507	∞	6.753	∞	**2.111**
	Sin	∞	∞	∞	5.919	∞	**6.334**	∞	**2.069**
	Uni	∞	∞	∞	5.746	∞	**6.211**	∞	**2.057**
	Tri	6.507	5.919	5.746	**2.566**	**7.291**	**3.322**	29.635	**2.087**
$\kappa_{04}^x = 0$	Gau	∞	∞	∞	**7.291**	∞	**7.311**	∞	**2.167**
$\kappa_{04}^x > 0$	Ray	6.753	**6.334**	**6.211**	**3.322**	**7.311**	**3.842**	23.212	**2.281**
	Lap	∞	∞	∞	29.635	∞	23.212	∞	3.750
	Exp	**2.111**	**2.069**	**2.057**	**2.087**	**2.167**	**2.281**	3.750	3.500

Table 8.2: Asymptotic values of $T\sigma^2_{\text{TOBSE}}$, in squared radians, for various source distributions. Values are computed via analytical expression (8.30), and the population statistics of Tables A.1 and A.2. Figures in boldface indicate source combinations for which the TOBSE provides better performance than the EML and the AEML estimators (cf. Table 5.1).

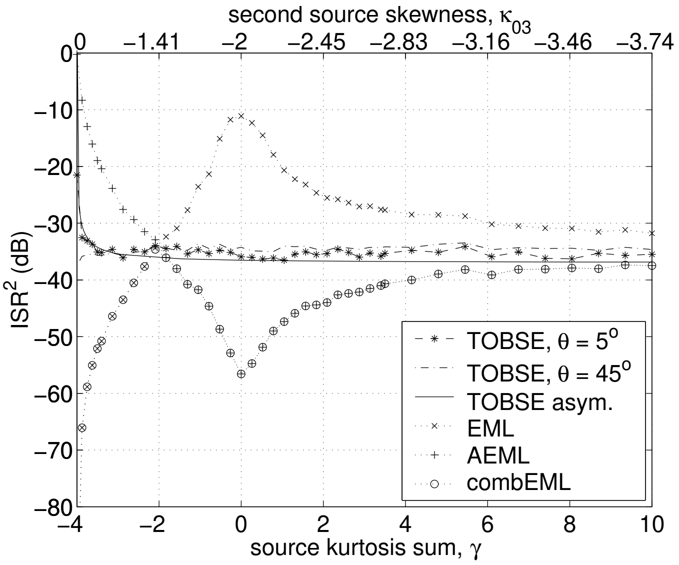


Figure 8.3: Performance of TOBSE (for $\theta = 5^o$ and $\theta = 45^o$), and 4th-order estimators EML, AEML and combEML (for $\theta = 5^o$) as a function of the source skewness and kurtosis sum. $T = 5 \cdot 10^3$ samples/signal, 200 MC iterations. The solid line shows theoretic asymptotic variance (8.30) ($10 \log_{10} \sigma^2_{\mathrm{TOBSE}}$).

structure of matrix Q helps to achieve this objective. The interest in solving the complex case stems from important applications such as multi-user digital communications [52] and seismic sounding [83], among others, where this type of mixtures naturally arises.

In the first place, Section 8.3.1 deals with the peculiarities of the problem in the complex case, and sets the statistical nomenclature that will be used for complex variables. Section 8.3.2 provides the definition and main properties of the *bicomplex numbers*, intimately related to the matrix group to which the transformation to be identified after pre-whitening belongs. By using this new tool, a family of closed-form estimators based on the HOS (the higher-order cumulants) of the whitened signals is disclosed in Section 8.3.3. This family is totally analogous to the one found for the real case in Section 8.2.1. At fourth order, the general expression reduces to the complex extension of the EML estimator. Similarly as occurred for its real equivalent, this 4th-order estimator shows a performance degradation when the sks is near zero. To overcome this deficiency, an alternative expression also based on the 4th-order cumulants is used, together with a decision rule to choose between both estimation expressions. The simulations reported in Section 8.3.5 illustrate these results and compare the proposed methods to other BSS techniques.

8.3.1 The complex-signal problem

Problem parameterization and equivalence class of valid solutions

In the complex case, the 2×2 unitary transformation Q becomes a unitary matrix, with the generic shape:

$$Q = \begin{bmatrix} a & -b^* \\ b & a^* \end{bmatrix}, \tag{8.34}$$

where $a,\, b \in \mathbb{C}$, such that $|a|^2 + |b|^2 = 1$, and symbol * represents the complex conjugation. Up to (irrelevant, due to the source scaling indeterminacy) pre-multiplication by a unit-norm diagonal matrix, the above matrix reduces to a complex Givens rotation, which exhibits the general form:

$$Q = \mathcal{Q}(\theta,\, \alpha), \tag{8.35}$$

with

$$\mathcal{Q}(\cdot,\,\cdot):\ \mathbb{R}\times\mathbb{R} \mapsto \mathbb{C}^{2\times 2}, \qquad \mathcal{Q}(\theta,\, \alpha) = \begin{bmatrix} \cos\theta & -\mathrm{e}^{-j\alpha}\sin\theta \\ \mathrm{e}^{j\alpha}\sin\theta & \cos\theta \end{bmatrix}. \tag{8.36}$$

Therefore, two parameters, angle θ and phase α, need to be estimated in order to achieve the identification of the relevant matrix, and hence the extraction of the sources. If $\hat{\theta}$ and $\hat{\alpha}$ represent estimates of θ and α, respectively, then the source separation is attained through

$$\hat{\mathbf{x}} = \widehat{Q}^{\mathrm{H}}\mathbf{z} = \widehat{Q}^{\mathrm{H}}Q\mathbf{x}, \qquad \widehat{Q} = \mathcal{Q}(\hat{\theta},\, \hat{\alpha}), \tag{8.37}$$

in which $^{\mathrm{H}}$ denotes the hermitian or complex-conjugate transpose operator. Observe that the actual values of $(\theta,\, \alpha)$ are not the only solutions to provide a valid source extraction. Due to the indeterminacies commented in Section 2.1.4, any solution of the form $\widehat{Q}^{\mathrm{H}}Q = PD$, with P a permutation and D a regular diagonal matrix is considered as acceptable, since it preserves the source waveforms. It is easily checked that the following set of solutions are equivalent in the waveform-preserving sense to $(\theta,\,\alpha)$, i.e., $(\hat\theta,\,\hat\alpha)\equiv(\theta,\,\alpha)$:

$$(\hat\theta,\,\hat\alpha)=\Big(\theta+\frac{m\pi}{2},\,\alpha+2\pi n\Big),\qquad \text{and} \tag{8.38a}$$
$$(\hat\theta,\,\hat\alpha)=\Big(-\theta+\frac{m\pi}{2},\,\alpha+(2n+1)\pi\Big),\qquad n,\,m\in\mathbb{Z}. \tag{8.38b}$$

By analogy with the real case (Section 3.2.1), we call

$$[(\theta,\,\alpha)]\triangleq\big\{(\hat\theta,\,\hat\alpha)\in\mathbb{R}^2\mid(\hat\theta,\,\hat\alpha)\equiv(\theta,\,\alpha)\big\} \tag{8.39}$$

the *equivalence class of valid solutions* of our complex problem.

Complex-variate statistics

The real-mixture methods studied up to now in this thesis rely on the higher-order cumulants of the data, and so do the complex versions considered next. There are different possible natural ways to define the cumulants for complex variables [5]. For our purposes, however, it is convenient to choose:

$$\mathrm{Cum}^z_{i_1 i_2 i_3\ldots}\triangleq\mathrm{Cum}[z^*_{i_1},\,z_{i_2},\,z^*_{i_3},\,\ldots]. \tag{8.40}$$

The same pairwise conventions as those defined in Section 2.1.2 hold (Kendall's convention [120]):

$$\kappa^z_{n-r,\,r}\triangleq\mathrm{Cum}^z_{\underbrace{1\ldots1}_{n-r}\underbrace{2\ldots2}_{r}}. \tag{8.41}$$

For instance, at 4th-order:

$$\begin{aligned}\kappa^z_{40}&=\mathrm{Cum}^z_{1111}=\mathrm{Cum}[z_1^*,\,z_1,\,z_1^*,\,z_1], &\quad \kappa^z_{31}&=\mathrm{Cum}^z_{1112}=\mathrm{Cum}[z_1^*,\,z_1,\,z_1^*,\,z_2]\\ \kappa^z_{22}&=\mathrm{Cum}^z_{1122}=\mathrm{Cum}[z_1^*,\,z_1,\,z_2^*,\,z_2], &\quad \kappa^z_{13}&=\mathrm{Cum}^z_{1222}=\mathrm{Cum}[z_1^*,\,z_2,\,z_2^*,\,z_2]\end{aligned}$$
$$\kappa^z_{04}=\mathrm{Cum}^z_{2222}=\mathrm{Cum}[z_2^*,\,z_2,\,z_2^*,\,z_2], \tag{8.42}$$

with similar definitions for other variables. Note that, according to these definitions, now certain cumulants may take complex values (such as κ^z_{31} and κ^z_{13} above). It be will seen that it is precisely from these complex-valued cumulants that the complex phase α will be estimated. For unit-variance zero-mean uncorrelated components, the expressions for the 4th-order cumulants become:

$$\begin{gathered}\kappa^z_{40} = \mathrm{E}\big[|z_1|^4\big] - \big|\mathrm{E}[z_1^2]\big|^2 - 2, \qquad \kappa^z_{31} = \mathrm{E}\big[z_1^*|z_1|^2 z_2\big] \\ \kappa^z_{22} = \mathrm{E}\big[|z_1|^2|z_2|^2\big] - 1, \qquad \kappa^z_{13} = \mathrm{E}\big[z_1^*|z_2|^2 z_2\big] \\ \kappa^z_{04} = \mathrm{E}\big[|z_2|^4\big] - \big|\mathrm{E}[z_2^2]\big|^2 - 2.\end{gathered} \tag{8.43}$$

Owing to the linear relationship between the whitened observations and the source signals represented by eqns. (3.1) and (8.35)–(8.36), and the multilinearity property of cumulants (Appendix A.2), we have:

$$\begin{aligned}\kappa^z_{40} &= \cos^4\theta\,\kappa^x_{40} + \sin^4\theta\,\kappa^x_{04} \\ \kappa^z_{31} &= \cos^3\theta\,\sin\theta\,\mathrm{e}^{j\alpha}\kappa^x_{40} - \sin^3\theta\,\cos\theta\,\mathrm{e}^{j\alpha}\kappa^x_{04} \\ \kappa^z_{22} &= \cos^2\theta\,\sin^2\theta\,(\kappa^x_{40} + \kappa^x_{04}) \\ \kappa^z_{13} &= \cos\theta\,\sin^3\theta\,\mathrm{e}^{j\alpha}\kappa^x_{40} - \sin\theta\,\cos^3\theta\,\mathrm{e}^{j\alpha}\kappa^x_{04} \\ \kappa^z_{04} &= \sin^4\theta\,\kappa^x_{40} + \cos^4\theta\,\kappa^x_{04},\end{aligned} \tag{8.44}$$

which are the complex counterparts of eqns. (3.11).

8.3.2 Bicomplex numbers

The specific nature of the matrix to be identified allows the development of a family of closed-form estimators analogous to the family proposed for the real case in Section 8.2.1, where the identification of the real orthogonal matrix required recourse to the complex numbers. Since complex-valued mixtures are now dealt with, we are naturally led to seek certain type of extension of the complex numbers. We carry out such extension as follows.

Definition 8.4 (bicomplex number). A *bicomplex number* $\bar{x} \in \mathcal{B}$ is an expression of the form:

$$\bar{x} = a + \jmath b, \qquad a, b \in \mathbb{C}, \quad \jmath^2 \triangleq -1. \tag{8.45}$$

Although symbol $\jmath$ has the same numeric value as the usual imaginary unit j, it must be understood as 'orthogonal' to or in a 'different space' from that of j, so that they actually represent distinct algebraic elements.

Definition 8.5 (breal and bimaginary parts). Given $\bar{x} = a + \jmath b \in \mathcal{B}$, $\mathbb{R}\mathrm{e}(\bar{x}) = a \in \mathbb{C}$ is the *bicomplex real (breal)* part of $\bar{x}$, and $\mathbb{I}\mathrm{m}(\bar{x}) = b \in \mathbb{C}$ its *bicomplex imaginary (bimaginary)* part.

This terminology is intended to avoid confusion with the commonplace real [$\mathrm{Re}(\cdot)$] and imaginary [$\mathrm{Im}(\cdot)$] parts of complex numbers, which are always real valued. Accordingly, $\jmath$ is named *bimaginary unit.*

Definition 8.6 (unitary bicomplex number). $\mathcal{B}_u = \{\bar{x} \in \mathcal{B} \ : \ |a|^2 + |b|^2 = 1\}$ is the set of *unitary* bicomplex numbers.

It is straightforward to establish an isomorphism between the set of bicomplex numbers and a particular set of matrices related to our problem. Let us first define those matrix sets.

Definition 8.7. $\mathcal{O}$ is the set of 2×2 *orthogonal* matrices with entries in $\mathbb{C}$, that is, $\mathcal{O} = \{U \in \mathbb{C}^{2\times 2} \ : \text{ all off-diagonal entries of } UU^{\mathrm{H}} \text{ and } U^{\mathrm{H}}U \text{ are zero}\}$.

Definition 8.8. $\mathcal{O}_u \subset \mathcal{O}$ is the set of *unitary* (or *orthonormal*) 2×2 matrices: $\mathcal{O}_u = \{U \in \mathbb{C}^{2\times 2} \ : \ UU^{\mathrm{H}} = U^{\mathrm{H}}U = I_2\}$.

The set $\mathcal{O}$ under ordinary matrix multiplication forms a non-Abelian group, as is simple to check. Accordingly, the set $\mathcal{O}_u$ under ordinary matrix multiplication is another non-commutative group, and more specifically a subgroup of SL(2, $\mathbb{C}$), the so-called special linear group of 2×2 matrices over $\mathbb{C}$ [63]. Any matrix $U \in \mathcal{O}$ can be written in general form as:

$$U = \begin{bmatrix} a & -b^* \\ b & a^* \end{bmatrix}, \qquad a, b \in \mathbb{C}. \tag{8.46}$$

The product of two orthogonal matrices is then given by:

$$U_1 U_2 = \begin{bmatrix} a_1 a_2 - b_1^* b_2 & -a_1 b_2^* - b_1^* a_2^* \\ b_1 a_2 + a_1^* b_2 & -b_1 b_2^* + a_1^* a_2^* \end{bmatrix}. \tag{8.47}$$

Now, we can define the *product* of two bicomplex numbers $\bar{x}_1$ and $\bar{x}_2$ in compliance with the product of $\mathcal{O}$-matrices by looking at the first column of the previous expression:

$$\bar{x}_1 \bar{x}_2 \triangleq \left[\mathbb{R}\mathrm{e}(\bar{x}_1)\mathbb{R}\mathrm{e}(\bar{x}_2) - \mathbb{I}\mathrm{m}^*(\bar{x}_1)\mathbb{I}\mathrm{m}(\bar{x}_2)\right] + \jmath\left[\mathbb{I}\mathrm{m}(\bar{x}_1)\mathbb{R}\mathrm{e}(\bar{x}_2) + \mathbb{R}\mathrm{e}^*(\bar{x}_1)\mathbb{I}\mathrm{m}(\bar{x}_2)\right]. \tag{8.48}$$

With this product operation: $\jmath^2 = -1$, in agreement with Definition 8.4. Also, if both the breal and bimaginary parts of the bicomplex operands are real, such product reduces to the usual complex-number multiplication (changing j for $\jmath$). Hence, eqn. (8.48) is a natural generalization of the latter. According to these conventions, the mapping

$$\aleph:\ \mathcal{O} \mapsto \mathcal{B}, \tag{8.49}$$

with

$$\aleph(U) = u_{11} + \jmath u_{21} = a + \jmath b, \qquad U \in \mathcal{O}, \tag{8.50}$$

represents an isomorphism between $\mathcal{O}$ under normal matrix multiplication and $\mathcal{B}$ under the bicomplex-number product as defined in (8.48). Effectively, $\aleph(\cdot)$ is one-to-one, onto and operation preserving, conditions that must be fulfilled by any isomorphism [63]. From the former two conditions, any $\bar{x} = a + \jmath b \in \mathcal{B}$ can be uniquely associated to the first column of one and only one matrix $U = [\mathbf{u}_1,\ \mathbf{u}_2] = \aleph^{-1}(\bar{x}) \in \mathcal{O}$, $\mathbf{u}_1 = [a,\ b]^{\mathrm{T}}$. The latter condition is due to (8.48), which guarantees $\aleph(U_1U_2) = \aleph(U_1)\aleph(U_2)$. Therefore, the related sets are isomorphic: $\mathcal{O} \simeq \mathcal{B}$ (and, similarly, $\mathcal{O}_u \simeq \mathcal{B}_u$). Owing to the properties of isomorphisms [63], the following definitions naturally follow, $\forall \bar{x} \in \mathcal{B}$:

Definition 8.9 (conjugation). $\bar{x}^* = \mathbb{R}\mathrm{e}(\bar{x})^* - \jmath\,\mathbb{I}\mathrm{m}(\bar{x})$ is the *bicomplex conjugate* of $\bar{x}$.

This is because the first column of U^{H} is $[a^*,\ -b]^{\mathrm{T}}$.

Definition 8.10 (modulus). $|\bar{x}|^2 = \bar{x}\bar{x}^* = \bar{x}^*\bar{x} = |\mathbb{R}\mathrm{e}(\bar{x})|^2 + |\mathbb{I}\mathrm{m}(\bar{x})|^2$.

Effectively, the first column of UU^{H} and $U^{\mathrm{H}}U$ is given by $\big[|a|^2 + |b|^2,\ 0\big]^{\mathrm{T}}$. As an immediate consequence:

Definition 8.11 (inverse). $\bar{x}^{-1} = \bar{x}^*/|\bar{x}|^2$.

For instance, $\jmath^{-1} = -\jmath$, just like its complex twin j.

If the unitary transformation is parameterized like in (8.35)–(8.36), then the associated unitary bicomplex number becomes an important particular case.

Definition 8.12 (bicomplex exponential). Expression

$$\mathrm{e}_{\alpha}^{\jmath\theta} \triangleq \cos\theta + \jmath \mathrm{e}^{j\alpha}\sin\theta, \qquad \theta,\ \alpha \in \mathbb{R}, \tag{8.51}$$

is called *bicomplex exponential*, by analogy with its familiar complex counterpart.

Thanks to the isomorphism established above, it is simple to prove the properties:

$$\text{(i)}\quad (\mathrm{e}_{\alpha}^{\jmath\theta})^n = \mathrm{e}_{\alpha}^{\jmath n\theta} \tag{8.52a}$$

$$\text{(ii)}\quad (\mathrm{e}_{\alpha}^{\jmath\theta})^{-1} = (\mathrm{e}_{\alpha}^{\jmath\theta})^* = \mathrm{e}_{\alpha}^{-\jmath\theta} \tag{8.52b}$$

$$\text{(iii)}\quad \mathrm{e}_{\alpha\pm\pi}^{\jmath\theta} = \mathrm{e}_{\alpha}^{-\jmath\theta} \tag{8.52c}$$

$$\text{(iv)}\quad \mathrm{e}_{\alpha}^{\jmath\theta_1}\mathrm{e}_{\alpha}^{\jmath\theta_2} = \mathrm{e}_{\alpha}^{\jmath\theta_2}\mathrm{e}_{\alpha}^{\jmath\theta_1} = \mathrm{e}_{\alpha}^{\jmath(\theta_1+\theta_2)} \tag{8.52d}$$

$$\text{(v)}\quad \mathrm{e}_{\alpha}^{\jmath(\theta+\frac{n\pi}{2})} = \begin{cases} (-1)^{\frac{n}{2}}\mathrm{e}_{\alpha}^{\jmath\theta}, & n \text{ even} \\ (-1)^{\frac{n-1}{2}}\jmath\,\mathrm{e}^{j\alpha}\mathrm{e}_{\alpha}^{\jmath\theta}, & n \text{ odd.} \end{cases} \tag{8.52e}$$

Observe that properties (8.52c) and (8.52e) may be interpreted as a compact reformulation in the bicomplex-number domain of the valid-solution equivalent class $[(\theta,\, \alpha)]$ stated in eqns. (8.38)–(8.39). The crucial point here [property (8.52e)] is that $\aleph^{-1}(\mathrm{e}_{\alpha}^{\jmath\frac{n\pi}{2}}) = PD$, $\forall n \in \mathbb{Z}$, P being a permutation and D a diagonal matrix with unit-norm diagonal elements, which, as commented in Section 2.1.4, do not affect the source-waveform recovery. Note, finally, that the bimaginary unit can be written as $\jmath = \mathrm{e}_0^{\jmath\frac{\pi}{2}}$, so that, as an important especial case of (8.52e):

$$\text{(vi)}\quad \jmath^n = \begin{cases} (-1)^{\frac{n}{2}}, & n \text{ even} \\ (-1)^{\frac{n-1}{2}}\jmath, & n \text{ odd,} \end{cases} \tag{8.52f}$$

similarly as occurs for the imaginary unit j.

The above results along with some straightforward algebraic manipulations easily show that:

Proposition 8.13 (bicomplex-number decomposition). *Any $\bar{x} \in \mathcal{B}$ can be decomposed as*

$$\bar{x} = \mathrm{e}^{j\omega}|\bar{x}|\,\mathrm{e}_{\varphi}^{\jmath\psi}, \tag{8.53}$$

with $\mathrm{e}^{j\omega} = \pm\mathrm{e}^{j\angle\mathbb{R}e(\bar{x})}$.

In order to obtain the values of the angle ψ and the phase φ, an extension of the argument (or angle) function is needed, the *bicomplex argument* or *"barg"* function, given below.

Definition 8.14 (barg function). The pair (ψ, φ) in decomposition (8.53) can be obtained through:

$$(\psi, \varphi) = \text{barg}(\bar{x}), \qquad \begin{cases} \psi = \text{arctg}\left(\left|\frac{\mathbb{Im}(\bar{x})}{\mathbb{Re}(\bar{x})}\right|\right) \\ \varphi = \arg\left(\frac{\mathbb{Im}(\bar{x})}{\mathbb{Re}(\bar{x})}\right), \end{cases} \tag{8.54}$$

where "arg$(\cdot)$" provides the angle (or argument) with respect the real axis, in $]-\pi, \pi]$, of its complex argument.

Bearing in mind properties (8.52), it is easy to see that if $\bar{x} = \mathrm{e}^{j\omega}|\bar{x}|\,\mathrm{e}_\alpha^{\jmath\theta}$ then the calculations of (8.54) lead to $\mathrm{e}_\varphi^{\jmath\psi} = \pm\mathrm{e}_\alpha^{\jmath\theta}$, where the sign depends on the value of θ. Specifically, if $\theta \in [-\pi/2, \pi/2]$ then $\mathrm{e}_\varphi^{\jmath\psi} = \mathrm{e}_\alpha^{\jmath\theta}$, while $\mathrm{e}_\varphi^{\jmath\psi} = -\mathrm{e}_\alpha^{\jmath\theta}$ for θ outside that range. As will be seen, this sign indeterminacy will not suffice to estimate the unknown parameters at 4th-order, so that the barg function will have to be redefined for such purposes (see Section 8.3.3).

In the real-mixture case, the so-called *scatter diagram* allows to simplify the development and provides an insightful geometrical interpretation of the real rotation taking place over the source signals after pre-whitening (see Sections 3.2.3, 3.3.4 and 5.2.2). If $\mathbf{z} = \mathcal{Q}(\theta)\,\mathbf{x}$, with $\mathcal{Q}(\theta) = \mathcal{Q}(\theta, 0)$ and $\mathbf{z}, \mathbf{x} \in \mathbb{R}^2$, then $z_1 + jz_2 = \mathrm{e}^{j\theta}(x_1 + jx_2)$ [eqn. (3.14)]. Taking into account bicomplex multiplication rule (8.48) and exponential expression (8.51), it is straightforward to prove that a similar relationship holds in the complex-mixture case too.

Proposition 8.15 (bicomplex scatter-diagram). *If* $\mathbf{z} = \mathcal{Q}(\theta, \alpha)\,\mathbf{x}$, *then*

$$z_1 + \jmath z_2 = \mathrm{e}_\alpha^{\jmath\theta}(x_1 + \jmath x_2). \tag{8.55}$$

By analogy to the real case, the bicomplex numbers $(z_1 + \jmath z_2)$ *and* $(x_1 + \jmath x_2)$ *are called* bicomplex scatter-diagram points *of the whitened output and the source signals, respectively.*

In the real case, the remaining orthogonal transformation after pre-whitening is easily interpreted as a geometric rotation of the scatter plots (Section 3.2.3). By contrast, in the complex case both the scatter-diagram points and the associated 'rotation' defined above, though algebraically analogous to the real-mixture ones, become less illuminating as a geometric notion.

Many of the above new concepts will allow us to simplify the notation of the complex-case closed-form estimators which are presented next. The beauty of the results on bicomplex numbers exposed in this section is that by constraining the breal and bimaginary parts to be real quantities, i.e., the associated transformation to have real elements, one is at once dealing with the familiar complex numbers and, in the context of BSS, with the real-mixture scenario.

8.3.3 Complex analytic estimators

Theorem 8.16. *Define $\xi_n(\mathbf{z})$ as the following bicomplex weighted sum of pairwise nth-order cumulants of the components of $\mathbf{z}$, with $n \in \mathbb{N}^+$:*

$$\xi_n(\mathbf{z}) \triangleq \sum_{r=0}^{n} \binom{n}{r} \jmath^r \kappa^z_{n-r,r}. \tag{8.56}$$

If $\mathbf{z} = \mathcal{Q}(\theta, \alpha)\,\mathbf{x}$, with $\mathbf{x}$ made up of independent components, then

$$\xi_n(\mathbf{z}) = \mathrm{e}_{\alpha}^{\jmath n\theta} \xi_n(\mathbf{x}), \quad \textit{for } n \textit{ even}, \tag{8.57}$$

where, according to (8.56),

$$\xi_n(\mathbf{x}) = \kappa^x_{n0} + \jmath^n \kappa^x_{0n}, \tag{8.58}$$

and

$$\xi_n(\mathbf{z}) = \mathrm{e}_{-\alpha}^{\jmath n\theta} \kappa^x_{n0} + \jmath^n \mathrm{e}_{\alpha}^{\jmath n\theta} \kappa^x_{0n}, \quad \textit{for } n \textit{ odd}. \tag{8.59}$$

Proof. The steps of the proof are just like those of Theorem 8.1, but now we have to take into account the complex conjugations in the definition of the cumulants [eqn. (8.40)]. Two cases must be distinguished. If n is even:

$$\kappa^z_{n-r,r} = \sum_{i_1 \ldots i_n} \underbrace{q^*_{1i_1} q_{1i_2} q^*_{1i_3} \cdots}_{n-r} \underbrace{\cdots q_{2i_{n-2}} q^*_{2i_{n-1}} q_{2i_n}}_{r} \mathrm{Cum}^x_{i_1 \ldots i_n}, \tag{8.60}$$

which, taking into account the source independence and that $q_{11} = \cos\theta$, $q_{12} = -\mathrm{e}^{-j\alpha}\sin\theta$, $q_{21} = \mathrm{e}^{j\alpha}\sin\theta$ and $q_{22} = \cos\theta$, develops into:

$$\kappa^z_{n-r,r} = (\cos\theta)^{n-r} \sin^r\theta \, \mathrm{e}^{j\alpha(r \bmod 2)} \kappa^x_{n0} + (-\sin\theta)^{n-r} \cos^r\theta \, \mathrm{e}^{j\alpha((r-n) \bmod 2)} \kappa^x_{0n}. \tag{8.61}$$

Since $\jmath^r = \jmath^n \jmath^{-(n-r)} = \jmath^n(-\jmath)^{n-r}$, (8.56) becomes:

$$\xi_n(\mathbf{z}) = \sum_{r=0}^{n} \binom{n}{r} \jmath^r \Big[(\cos\theta)^{n-r} \sin^r\theta \, \mathrm{e}^{j\alpha\,(r \bmod 2)} \kappa^x_{n0} + \\ + (-\sin\theta)^{n-r} \cos^r\theta \, \mathrm{e}^{j\alpha\,((n-r)\bmod 2)} \kappa^x_{04} \Big] = \\ = \kappa^x_{n0} \sum_{r=0}^{n} \binom{n}{r} (\cos\theta)^{n-r} (\jmath \sin\theta)^r \mathrm{e}^{j\alpha\,(r\bmod 2)} + \\ + \jmath^n \kappa^x_{0n} \sum_{r=0}^{n} \binom{n}{r} (\jmath \sin\theta)^{n-r} \cos^r\theta \, \mathrm{e}^{j\alpha\,((n-r)\bmod 2)}. \quad (8.62)$$

Now, the breal part of the first summation is

$$\binom{n}{0} \cos^n\theta - \binom{n}{2} (\cos\theta)^{n-2} \sin^2\theta + \cdots = \cos(n\theta), \quad (8.63)$$

whereas the bimaginary part reads

$$\left[\binom{n}{1} (\cos\theta)^{n-1} \sin\theta - \binom{n}{3} (\cos\theta)^{n-3} \sin^3\theta + \dots \right] \mathrm{e}^{j\alpha} = \mathrm{e}^{j\alpha} \sin(n\theta). \quad (8.64)$$

The second sum produces exactly the same results, so:

$$\xi_n(\mathbf{z}) = \big(\cos(n\theta) + \jmath \mathrm{e}^{j\alpha} \sin(n\theta)\big)(\kappa^x_{n0} + \jmath^n \kappa^x_{0n}) = \mathrm{e}^{\jmath n\theta}_{\alpha} (\kappa^x_{n0} + \jmath^n \kappa^x_{0n}). \quad (8.65)$$

The first half of the theorem is then proven, since $\xi_n(\mathbf{x}) = \kappa^x_{n0} + \jmath^n \kappa^x_{0n}$. For n odd, however:

$$\kappa^z_{n-r,r} = \sum_{i_1 \dots i_n} \underbrace{q^*_{1i_1} q_{1i_2} q^*_{1i_3} \dots}_{n-r} \underbrace{\dots q^*_{2i_{n-2}} q_{2i_{n-1}} q^*_{2i_n}}_{r} \mathrm{Cum}^x_{i_1 \dots i_n}, \quad (8.66)$$

which turns into

$$\kappa^z_{n-r,r} = (\cos\theta)^{n-r} \sin^r\theta \, \mathrm{e}^{-j\alpha(r \bmod 2)} \kappa^x_{n0} + (-\sin\theta)^{n-r} \cos^r\theta \, \mathrm{e}^{j\alpha((n-r)\bmod 2)} \kappa^x_{0n}. \quad (8.67)$$

So now the bimaginary part of the first sum becomes $\mathrm{e}^{-j\alpha}\sin(n\theta)$, whereas the rest of the terms remain just as before. Hence, for n odd:

$$\xi_n(\mathbf{z}) = \big(\cos(n\theta) + \jmath \mathrm{e}^{-j\alpha} \sin(n\theta)\big)\kappa^x_{n0} + \jmath^n \big(\cos(n\theta) + \jmath \mathrm{e}^{j\alpha} \sin(n\theta)\big)\kappa^x_{0n} = \\ = \mathrm{e}^{\jmath n\theta}_{-\alpha} \kappa^x_{n0} + \jmath^n \mathrm{e}^{\jmath n\theta}_{\alpha} \kappa^x_{0n}. \quad \blacksquare$$

This theorem is a generalization to complex mixtures of Theorem 8.1. Yet the result for the complex case does not seem to lead to a simple estimation of the unknown parameter at odd orders [eqn. (8.59)]. In the real case, by contrast, a third-order estimator (TOBSE, see Section 8.2.1) was derived from the associated family. This is not a strong shortcoming, since in practice many complex signals of interest are circularly distributed processes, whose odd-order cumulants all vanish [5]. For even order, however, the situation is quite different. In that case, $\xi_n(\mathbf{x})$ is a real number, and then parameters $(\theta,\, \alpha)$ may be easily estimated from (8.57) via:

$$(n\hat{\theta},\, \hat{\alpha}) = \mathrm{barg}\left(\frac{\xi_n(\mathbf{z})}{\xi_n(\mathbf{x})}\right), \qquad n \text{ even}, \tag{8.68}$$

where "barg$(\cdot)$" represents the bicomplex angle function defined in (8.54).

Estimators (8.68) exhibit analogous indeterminacy problems as their real equivalents exposed in Section 8.2.1. This indeterminacy stems from the fact that $n\theta = n(\theta + 2\pi m/n)$, for any $n \in \mathbb{N}^+$, $m \in \mathbb{Z}$, so a non-valid separation solution may be produced. Also, in principle the value of $\xi_n(\mathbf{x})$ (i.e., the source marginal nth-order cumulants) would need be known a priori. But these two apparent difficulties do not exist at all orders.

Order 4

At fourth-order, (8.57) becomes:

$$\xi_4(\mathbf{z}) = (\kappa^z_{40} - 6\kappa^z_{22} + \kappa^z_{04}) + \jmath 4(\kappa^z_{31} - \kappa^z_{13}) = \mathrm{e}_\alpha^{\jmath 4\theta}(\kappa^x_{40} + \kappa^x_{04}) = \mathrm{e}_\alpha^{\jmath 4\theta}\gamma, \tag{8.69}$$

so that the unknown parameters can be obtained from

$$(4\hat{\theta},\, \hat{\alpha}) = \mathrm{barg}\big(\xi_4(\mathbf{z})/\xi_4(\mathbf{x})\big). \tag{8.70}$$

Due to the evident connection with its real counterpart, this is called the *Complex EML (CEML)* estimator. We might as well have obtained the same result by developing the 4th-order cumulants (8.44) directly. Effectively:

$$\begin{aligned} \kappa^z_{40} - 6\kappa^z_{22} + \kappa^z_{04} &= \cos(4\theta)\gamma \\ 4(\kappa^z_{31} - \kappa^z_{31}) &= \mathrm{e}^{j\alpha}\sin(4\theta)\gamma, \end{aligned} \tag{8.71}$$

Just like its real twin, the sks $\gamma = \xi_4(\mathbf{x})$ may be obtained from the available signals as in eqns. (3.26) and (3.27):

$$\gamma = \kappa^z_{40} + 2\kappa^z_{22} + \kappa^z_{04} = \kappa^x_{40} + \kappa^x_{04}. \tag{8.72}$$

The application of the barg function must be done here with some extra care. Property (8.52e) along with the remarks after Definition 8.14 prove that eqn. (8.54) leads to a $\pm 45^{\rm o}$ bias in the estimation of θ whenever $|\theta| > \pi/2$, because $-\mathrm{e}_{\alpha}^{\jmath 4\theta} = \mathrm{e}_{\alpha}^{\jmath(4\theta\pm\pi)}$. Similar problems arise in the AML estimator of [69] when employing the arctg function (Section 3.3.2). To get round this difficulty, we can take advantage of the fact that the breal part of $\xi_4(\mathbf{z})/\xi_4(\mathbf{x})$ is real, so that the "$\pm$-uncertainty" of decomposition (8.53) may thus be resolved through

$$\psi = \arg\big(\mathbb{R}\mathrm{e}(\bar{x}) + j|\mathbb{I}\mathrm{m}(\bar{x})|\big) \tag{8.73}$$

$$\varphi = \arg\big(\mathbb{I}\mathrm{m}(\bar{x})\big). \tag{8.74}$$

The above redefinition of the barg function prevents any bias, since, by bearing in mind property (8.52c), $\mathrm{e}_{\varphi}^{\jmath\psi} = \bar{x}$, $\forall \bar{x}$ of the form $\mathrm{e}_{\alpha}^{\jmath\theta}$.

We conclude this section with two final remarks. First, recall that in [41] relationship (3.38) was found for the whitened-observation 4th-order cumulants in the real case. Note that relating the norms of the terms at both sides of the second equality in (8.69), and using $|\rho|^2$ instead of ρ^2, one arrives at the complex equivalent of such cumulant relationship.

Second, in contrast to the real case, we have observed an apparent lack of compact scatter-plot based expressions for the bicomplex centroids. The reader may reach this conclusion by trying to compute, e.g., $(z_1 + \jmath z_2)^4$ using the bicomplex-number algebra developed in Section 8.3.2. Perhaps the bicomplex product may need to be specially redefined to make such scatter-plot expressions available in the complex case as well.

8.3.4 Combined estimation strategy

The complex AEML and combEML

As in the real case, the CEML estimator (8.70) cannot be used for zero sks, since then $\xi_4(\mathbf{x}) = 0$ (the performance degradation of this estimator around $\gamma = 0$ will be demonstrated in the experiments of Section 8.3.5). To overcome this deficiency, we resort to the complex version of the AEML estimator, put forward in Section 5.2. In the complex case, and by virtue of relationships (8.44), the associated cumulant combination develops into:

$$\xi_4'(\mathbf{z}) = (\kappa_{40}^z - \kappa_{04}^z) + 2\jmath(\kappa_{31}^z + \kappa_{13}^z) = \mathrm{e}_\alpha^{\jmath 2\theta}\eta, \tag{8.75}$$

where, $\eta = \kappa_{40}^x - \kappa_{04}^x$ represents the skd, whose actual sign is unimportant for the estimation of the true parameters [due to equivalence class (8.38) and property (8.52e)]. Then,

$$(2\hat{\theta}, \hat{\alpha}) = \mathrm{barg}\big(\xi_4'(\mathbf{z})\big). \tag{8.76}$$

For the sake of consistency in the nomenclature, we name eqn. (8.76) the *Complex AEML (CAEML)* estimator.

The identical issue as in the real case about when to use (8.70) or (8.76) naturally arises at this point. In a bid to shed some light on this issue, experiments along the lines of those described in Section 5.3.3 were carried out with the complex versions. This time, 4-QAM signals were used as sources, with independent real and imaginary parts composed of respective real-valued PRBS's with the same symbol probability. In these conditions, the (normalized) kurtosis of the resultant 4-QAM signal is $\kappa_4^x = \kappa_{4\,(\mathrm{PRBS})}^x/2$, so that any kurtosis value equal or greater than -1 can be obtained by suitably varying the PRBS symbol probability (these results are exposed in Appendix B). Signals were composed of $T = 5000$ samples, and performance measure ISR^2 (Section 3.3.7) was computed over 100 Monte Carlo runs. The results obtained, plotted in Figure 8.4, were very similar to those of Section 5.3.3. The alternation in best performance occurs when the skd is equal to the sks (around $\gamma = \eta = -1$), with the CEML offering lower ISR for $|\gamma| > |\eta|$, and vice versa. On the other hand, from eqns. (8.69) and (8.75), and Defs. 8.10 and 8.12, it turns out $|\xi_4(\mathbf{z})| = |\gamma|$ and $|\xi_4'(\mathbf{z})| = |\eta|$ (just like for their real counterparts). This leads to the conclusion that decision rule

$$|\xi_4(\mathbf{z})| \underset{\mathrm{CAEML}}{\overset{\mathrm{CEML}}{\gtrless}} |\xi_4'(\mathbf{z})| \tag{8.77}$$

may as well be applied in the complex case. Although $\theta = 30^\mathrm{o}$ and $\alpha = 45^\mathrm{o}$ were used in Fig. 8.4, this conclusion is essentially identical for all angle values (θ, α). Fig. 8.4 also shows the results that would have been obtained under the same conditions by using the combined estimation scheme derived from the above rule, that we refer to as *combCEML* estimator. The performance degradation exhibited by the CEML and the CAEML around $\gamma = 0$ and $\eta = 0$, respectively, is now corrected with the combCEML.

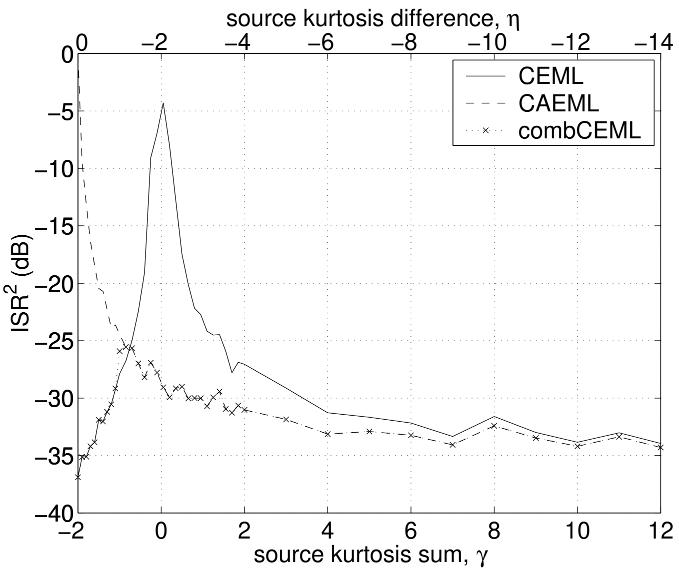


Figure 8.4: Performance of CEML, CAEML and combined CEML vs. sks, for $\theta = 30^o$ and $\alpha = 45^o$. Sample size $T = 5 \cdot 10^3$, 100 MC runs.

Connections

In [44,45] another two closed-form estimators are proposed to solve the complex-mixture problem from the pre-whitened signals. The first one (which is also featured in [53] and [80], and we refer to as "CCF", for "complex CF") is simply the complex extension of the real CF estimator developed in [41] and analyzed in Section 3.4. Interestingly, the second one happens to coincide[1] with the complex alternative estimator suggested in the previous section. To choose between the two, the decision rule [see (5.35)]

$$|\kappa_{22}^z| \underset{\text{2nd.}}{\overset{\text{1st.}}{\gtrless}} |\kappa_{31}^z + \kappa_{13}^z| \tag{8.78}$$

is utilized, the rationale behind it being the conditioning of the respective parameters ρ [like that in eqn. (3.38)] of the estimators. For consistency in the nomenclature, we refer to the combined estimator derived from this rule as "combCCF".

It could seem, therefore, that (8.76) has no merit, for an equivalent expression was already defined in a previous work [44,45]. Nonetheless, as occurred for the AEML (Section 5.2.1), the estimator is equipped here with a simple yet smart algebraic formalism in terms of 4th-order cumulant bicomplex centroids. This formalism is in accordance

[1] With the correction pointed out in Footnote 1, Chapter 5.

with the unified notation used for the other closed-form methods considered in this thesis, and it indeed proves beneficial, for it allows us: 1) to easily interrelate estimator (8.76) to the estimators studied in Sections 8.2.1 and 8.3.3, and 2) to obtain a practical decision rule for the combCEML combined estimation strategy, which, we anticipate, will prove more effective than the combCCF rule (8.78). This last claim is put to the test in the next section.

8.3.5 Simulation results

This section reports some numerical simulations that illustrate and endorse the preceding theoretical exposition. We also assess and compare the relevant methods.

Variation with angular parameters

We first evaluate estimators CEML (8.70), CAEML (8.76), and CCF [44,45]. We are primarily concerned with their behaviour as a function of angular parameters θ and α. Figure 8.5 plots the results for varying θ, with 16-QAM and complex Gaussian sources. It is seen how both the CCF and the CEML performance depends on θ, whereas this does not occur for the CAEML. Nevertheless, the performance of these methods is independent of parameter α, as illustrated by Figure 8.6.

Comparison among several complex BSS methods

In the second place, the combined estimation strategies based on decision rules (8.77) and (8.78) are assessed and compared to other methods. The other methods considered are CEML, SOK, SOK' and JADE. Recall that SOK is the closed-form solution given in [52] to the maximization of the contrast function (4.13) of [94]. It was shown in Chapter 4 that such solution corresponds to the EML when instead of the equal sign of source kurtosis ε the sign of sks ε_γ is used. In this manner, the applicability of such contrast function was expanded, since the need for a prior knowledge of the signs of source kurtosis and for these signs to be equal is spared. In the following experiments it was assumed for SOK $\varepsilon = 1$. SOK' corresponds to solution SOK, but using in the associated algorithm the sign of sks as calculated from the whitened data by (8.72).

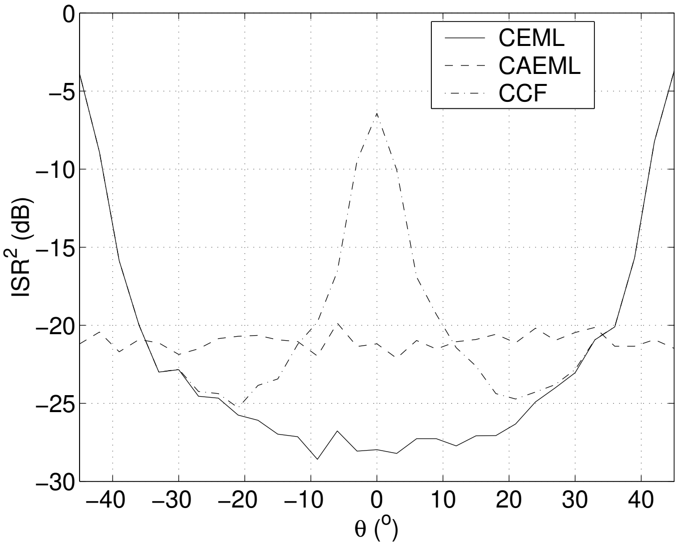


Figure 8.5: Mean squared ISR vs. parameter θ. Sources: 16-QAM and complex Gaussian. $\alpha = 65^o$. Sample size $T = 5 \cdot 10^3$, 100 MC runs.

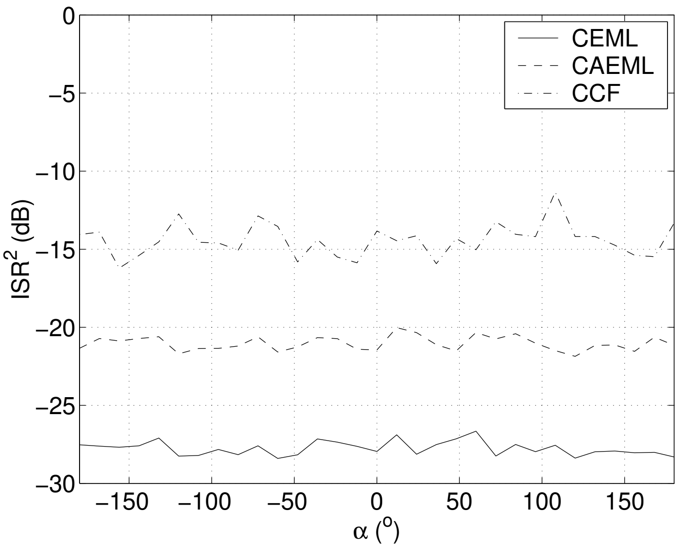


Figure 8.6: Mean squared ISR vs. parameter α. Sources: 16-QAM and complex Gaussian. $\theta = 5^o$. Sample size $T = 5 \cdot 10^3$, 100 MC runs.

In this comparison experiment, different sks are simulated by means of 4-QAM signals of appropriate symbol probability (as commented in Section 8.3.4; see also Appendix B). Figure 8.7 summarizes the results. Note, first, the close performance trends between the JADE and the combCEML, as was also evidenced in the real case (Section 5.3.4). On the other hand, decision rule (8.78) does not seem to be effective, since the combCCF performance considerably worsens as the kurtosis of the sources tend to be equal. Results for CEML and SOK' are exactly identical besides the region near $\gamma = 0$, around which the former improves the latter. This suggests that the conclusion drawn in Section 4.3 regarding the relationship between EML and SOK might perhaps be extended to the complex case as well, even though the complex SOK solution makes use of a whitened-observation 4th-order cross-cumulant which is not regarded by the CEML [52]. However, further experiments (Figure 8.8) confirm that, as was to be expected from the CEML performance variation with θ revealed by Fig. 8.5, this similarity vanishes for higher values of θ,[2] although the performance of both methods remain extremely close. Fig. 8.8 also seems to indicate that the performance of the combCCF gets closer to that of the combCEML for higher θ.

8.4 Summary and Conclusions

This chapter has been dedicated to extending a number of results on closed-form estimators for BSS in two fronts: statistical order and complex mixtures.

The general form of complex linear combinations of the whitened-signal nth-order cumulants such that they preserve explicit analytic expressions for the parameter of interest has been unveiled. These complex linear combinations generalize the concept of centroid, which was initially defined in connection with the 4th-order estimators studied earlier in this thesis. Optimization criteria from which the resulting estimators implicitly stem have also been obtained. A difficulty regarding the solution uniqueness has been encountered at orders greater than four. However, this has not prevented the general expression from being of practical benefit in some cases. At fourth order, the general form results in the EML estimator. At third order, a novel closed-form

[2]For 'high θ' is understood to have values around $(2n+1)\pi/4$, $n \in \mathbb{Z}$, since the equivalent class of valid solutions (8.38) is $\frac{\pi}{2}$-periodic in θ.

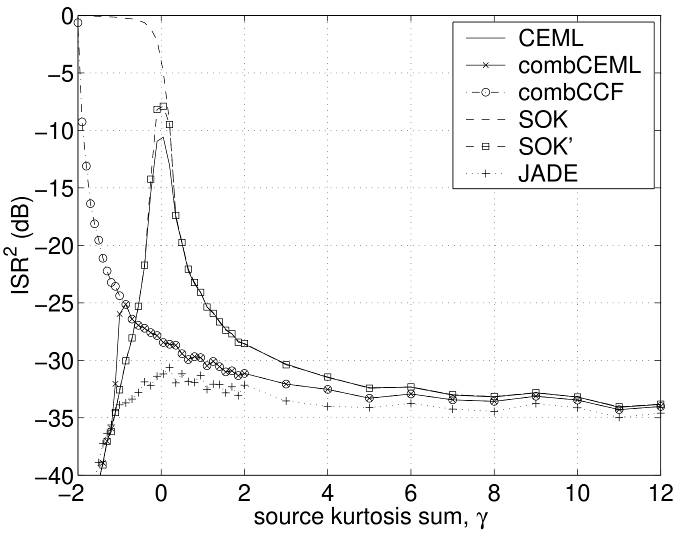


Figure 8.7: Mean squared ISR vs. sks for several complex-mixture BSS methods. $\theta = 5^o$ and $\alpha = 65^o$. Sample size $T = 5 \cdot 10^3$, 200 MC runs.

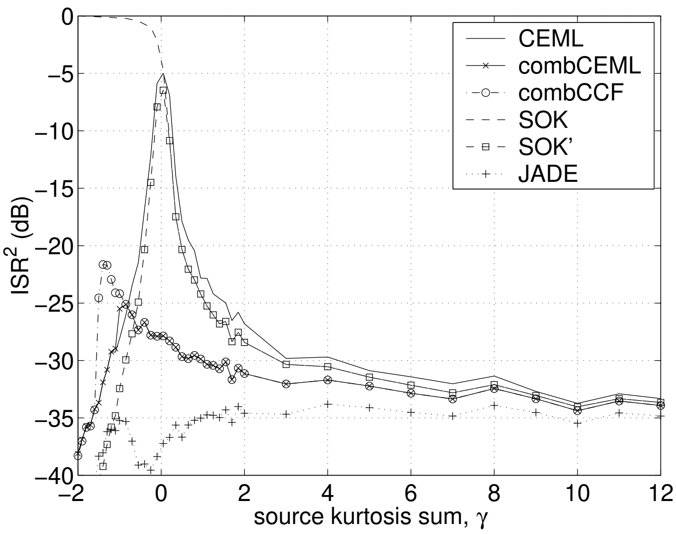


Figure 8.8: Mean square ISR vs. sks for several complex-mixture BSS methods. $\theta = 30^o$ and $\alpha = 65^o$. Sample size $T = 5 \cdot 10^3$, 200 MC runs.

estimator, the TOBSE, has been derived and analyzed along the same lines as the other estimators previously considered. Although it requires that at least one of the sources has an asymmetric distribution, the lower estimation errors and, in many cases, better asymptotic performance relative to other 4th-order methods (such as EML and AEML) make this new 3rd-order estimator worth considering when facing a real-mixture BSS problem.

On the other hand, we have defined the bicomplex numbers, forming a set isomorphic to the group of transformations relevant to the BSS problem after pre-whitening. These algebraic devices have allowed an elegant extension to the complex-mixture case of various closed-form estimation ideas in real scenarios. In fact, previous results on real mixtures constitute straightforward particular cases of the results on complex mixtures developed in this chapter, including the concepts of scatter diagram and centroid. The extended centroids arise as specific bicomplex linear combinations of the whitened-vector cumulants which are able to retain the information contained in the unitary mixing transformation, giving rise to compact closed-form expressions for the estimation of the pertinent parameters. Altogether, a unified formulation of the real and the complex problems within the framework of their analytical solutions has been reached.

Chapter 9

Conclusions and Outlook

9.1 Summary and Conclusions

This thesis has been devoted to the problem of blind separation of independent source signals from the observation of instantaneous linear mixtures of the sources. The two-stage approach to solving the problem has been adopted throughout. This approach is composed of a first SOS-based step — so-called pre-whitening — followed by a second step in which the identification of the remaining unitary mixture requires the exploitation of the HOS of the data when the time structure of the observed processes is ignored or it just cannot be utilized. In the two-signal case, the unitary transformation can be easily parameterized. This work has studied in detail existing procedures and put forward new procedures for the closed-form estimation of those parameters. We want to highlight the term *closed-form* estimation as opposed to iterative or non-direct (often optimization-based) approaches. The advantage of the closed-form approach is manifold but, in particular, it has very favourable implications in performance evaluation: analytic-formula based estimation has allowed us to obtain further closed-form expressions for the statistical characterization of the different methods.

Chapter 3 has explored an approximate maximum-likelihood (AML) estimator developed by Harroy and Lacoume [69]. Originally, the conditions of its development made it only valid under symmetric sources with the same distribution and certain kurtosis values. We have developed a similar formula — the *extended ML (EML)* —

which is valid for any source distribution, with the only condition that the source kurtosis sum (sks) not be zero. Contrasting both expressions, the EML has been seen as a generalization of the AML, for the former is also the ML estimator of the unknown parameter under the conditions of the latter. The new estimator has been endowed with a compact scatter-plot centroid based expression, whose use has been twofold. In the first place, it has allowed to easily understand the EML estimation operation in geometrical terms, when symmetric identically distributed sources are processed. These illuminating geometrical ideas have been inspired by the work of Bogner [13] and Clarke [39]. In the second place, the simple algebraic formalism has simplified the characterization of the statistical properties and the asymptotic performance analysis of the estimator. In fact, a common analysis framework has been developed, in terms of which many of the other estimators later considered have also been easily analyzed. In particular, a closed-form expression for the estimator asymptotic probability density function (pdf) has been determined under quite general conditions (essentially, i.i.d. mixtures), as well as a Gaussian approximation valid when the sample size is large enough. These expressions are able to predict the estimation performance from the specification of the source distributions. A variety of experiments have validated the theoretical results.

The second part of Chapter 3 has also considered another closed-form estimation expression, the so-called Comon's direct formula (CF). By means of the common analysis framework based on the complex-centroid formalism, this estimator has been understood in detail. Interestingly, its performance has been found to depend on the unknown parameter. This fundamental drawback makes the EML emerge as the best option among the above three estimators considered. All these theoretical results have been supported through various simulations.

Chapter 4 has found the optimization criteria that the EML estimation expression implicitly solves in closed-form. They are a cumulant-matching minimization criterion and a contrast-function maximization criterion, both intimately related to each other. The former has been shown to correspond to a peculiar optimal output cumulant-matching principle in which no prior knowledge of the source statistics is necessary. The latter has proved specially fruitful, for it has been shown to correspond to the extension of a contrast function by Moreau [94,95]. Moreau's contrast function (MCF) could only be applied to sources with the same (known in advance) kurtosis sign. Yet, by virtue of the EML contrast function, it is seen that in the two-signal case a simple modification makes it valid for any source kurtosis sign (with no need for prior knowledge), as long as the applicability conditions of the EML estimator hold. As a by-product of this interesting discovery, a rather involved analytic solution found by Comon-Moreau [52] for MCF has been shown to reduce, in the noiseless case and with the mentioned minor alterations, to the EML solution.

Chapter 5 has proposed another closed-form fourth-order estimator, the *alternative EML (AEML)*, mainly motivated by the performance degradation of the previous 4th-order methods for near zero sks. Even though the AEML already appears in the literature [44,45], here it has been reformulated within the coherent algebraic formalism of complex centroids, which has allowed us, among other benefits, to study its asymptotic performance borrowing the analysis tools used for the previous methods. According to this analysis, the AEML has been found to show a performance deterioration with the source kurtosis difference (skd). In an attempt to prevent this undesirable performance variation with the source statistics, the EML and the AEML have then been combined through a suitable decision rule to choose between the two estimators given a set of whitened observations. A suboptimal heuristic decision rule has been derived, based on empirical outcomes, and later shown to agree in many cases with the optimal, in the minimum mean square error sense, rule. Experiments have demonstrated a performance variation of the resulting combined estimator — so-called *combined EML (combEML)* — similar to the trends exhibited by contrast-based methods such as Comon's HOEVD [48] and Cardoso's JADE [23]. The combEML improves HOEVD and JADE in a number of source distribution combinations.

Chapter 6 has accomplished the extension of the methods to the general BSS scenario composed of more than two measurements of (possibly) more than two sources. The simple Jacobi-like pairwise iteration strategy originally proposed in [45, 48] has been adopted, providing satisfactory performance in virtually all cases. These extensions have been applied to a biomedical problem: the non-invasive extraction of the fetal electrocardiogram (ECG) signals from maternal corrupted cutaneous electrode recordings. Very successful results have been obtained on real ECG recordings taken from a pregnant woman, evidencing a level of performance up to the mark of other methods such as the HOEVD and JADE, and considerably superior to classical methods such as Widrow's multi-reference adaptive noise cancellation [129].

Chapter 7 has considered, for their application in non-stationary separation environments, the adaptive implementation of some of the methods studied previously, which are in nature block (or batch, or off-line) processing procedures. We have focused on the EML, the AEML and the combEML methods, giving rise to their adaptive counterparts, *adEML*, *adAEML* and *adcombEML*, respectively. It has been shown that implementations based on the adaptive update of the complex centroids present a number of advantages over versions based on the stochastic optimization of the associated contrast functions. In particular, a very fast convergence has been observed, whose speed is almost independent of the adaption coefficient. In addition, the centroid updating has shown the form of typical stochastic adaptive algorithms, so that the performance has been studied with the conventional tools for their analysis. Global convergence free of spurious solutions has been proven for these adaptive methods, as long as the conditions on the source statistics of their respective batch counterparts are fulfilled. Analytic expressions which accurately predict their asymptotic behaviour given the separation system parameters have been obtained as well. The pairwise extension to more than two signals has been implemented, offering a computational cost and a performance level comparable to other on-line procedures.

Finally, Chapter 8 has been dedicated to extending some of the previous results in two fronts. In the first place, cumulants of *any* order have been considered. A family of nth-order cumulant based complex centroids has been determined, which retain explicit expressions for the estimation of the unknown mixing parameter. The family has been associated with respective optimization criteria. At $n = 4$, this general class yields the EML method. In the case $n = 3$, a new third-order estimator has been derived, the *third-order blind signal estimator (TOBSE)*, applicable when at least one of the sources has an asymmetric distribution. Again, thanks to its centroid formulation, its statistical characterization has been easily studied. TOBSE's asymptotic performance has been shown to improve that of EML and AEML in a good number of cases. The lower estimation error of 3rd-order cumulants relative to 4th-order cumulants is another strong point of TOBSE.

The computational complexity of EML, CF, AEML and TOBSE has also been examined in their respective chapters, with two cases being distinguished depending on whether the centroids are calculated from the cumulants or from the equivalent scatter-plot based expressions. There has been no agreement among the different methods as to which of these two forms is less expensive, despite the more appealing higher notational compactness of the scatter-plot based expressions.

In the second place, Chapter 8 has also extended the general family of estimators, as well as other real-case results, to the complex-mixture case. To arrive at this extension, a new class of numbers — the so-called *bicomplex numbers* — has been defined taking into account the structure and properties of the unitary mixing transformation to be identified. The beauty in the use of the bicomplex number lies in that they have allowed a formulation of the closed-form solutions to the complex BSS problem within the common framework developed for the real-case analytical solutions, thus manifesting the existing tight bonds between real and complex scenarios in the closed-form solution context.

In conclusion, this thesis has provided the first unified global vision of the analytic estimators for BSS. Although only the instantaneous linear mixture case has been considered, it is envisaged that these results may find application in other scenarios, such as convolutive mixtures. It is hoped by the author that this work will open fresh new avenues of research on the exciting topic of BSS.

9.2 Outlook

The following is a list of possible points which could lead the continuation of the present investigation:

- Noise has been disregarded in this work, yet the noise impact on the performance of the methods studied herein is an important issue to be explored. Preliminary experimental results in noise environments reported in [137] and [141] show a good performance by EML, while revealing unexpected anomalous behaviour by HOEVD and JADE in particular separation scenarios. Other experiments with noise appear in [138].

- Finding out whether these results on closed-form estimators for the instantaneous linear case could be extended to the more general convolutive-mixture problem merits further work, for the convolutive case presents a good number of important applications in fields such as speech processing and mobile communications.

- Further connections between the optimization criteria unfolded in Chapter 4 and other contrast-based criteria in the literature (such as those of the HOEVD and JADE) could be sought. This would allow a better understanding of the existing relationships with those methods.

- The extension to more than two signals of the EML contrast function (4.14)–(4.16) could be sought.

- Alternative combined estimation strategies could be contemplated, with estimator pairs other than those considered herein, or where perhaps even more than two expressions were combined.

- The methods developed here could be applied to, and compared within, problems of other disciplines, such as some of those commented in Section 1.1.

- Even though the pairwise extension to more than two signals has always provided satisfactory empirical results, the convergence of the procedure still remains to be proved theoretically.

- Some type of phase-unwrapping strategy could be developed in order to surmount the indeterminacy problem of the general estimation family for $n > 4$ encountered in Chapter 8. Whether this would actually be beneficial should also be a subject of careful consideration, since the estimation error at such high orders could render any effort in that direction worthless.

- Only a few properties of the bicomplex numbers presented in Chapter 8 have been utilized in this thesis; however, it is envisaged that all their potential may be used to generate further results on this topic, or even on problems of other fields where they may prove useful, hence opening exciting new investigation areas.

– The asymptotic performance of the complex estimators needs to be theoretically analyzed. To accomplish this task, the analysis tools developed here for the real estimators could perhaps be exploited.

– Alternative extensions to more than two signals could be looked into as well. We propose two possible paths. The first would consider (in the real case) complex-centroid estimators (like those studied here) with cumulants involving more than two signals. The second alternative would be based in a natural extension of the complex numbers, the so-called *quaternions*, and it is explained in more detail next.

9.2.1 Blind separation via quaternions

We conclude this thesis with a promising interesting proposal: the application of quaternions to BSS.

Quaternions

Quaternions were invented by Sir William Rowan Hamilton, the most important Irish mathematician of all time, in the 1840's [67]. His original motivation was mostly geometrical, as he was trying to find out what the result of dividing a three-dimensional vector by another was. The answer was, precisely, a quaternion.

Algebraically, a quaternion is a four-dimensional entity which can be represented as a linear combination of the four quaternion units 1, i, j, and k. These units form the basis for the quaternion space, and fulfil the famous fundamental relations $i^2 = j^2 = k^2 = ijk = -1$, which give the basic rules for quaternion multiplication. While quaternions share many properties with complex numbers (as a matter of fact, they can be considered as a natural extension of these), they exhibit the remarkable feature that their product is not commutative [67, 77]. In fact, they constituted the first non-Abelian ring to be discovered [63].

One of the most attractive features of quaternions is their ability to represent and perform operations in the three-dimensional space, including affine transformations, projections and rotations. References [1, 57] provide a thorough account of these capabilities, with emphasis on the representation of rotations via quaternions.

Application of quaternions to BSS

Consider a three-signal BSS problem. As explained in Chapter 3, in the real case the whitened-observation jpdf would be a rotated version, in a three-dimensional space, of the independent-source jpdf. In the mentioned chapter we took advantage of the ability of complex numbers to deal with rotations in the two-dimensional space (represented by the complex plane) in order to find the unknown angular parameter in the two-signal scenario. Our idea is therefore to benefit from the potential of quaternions to perform rotations in three dimensions in order to find the unknown rotation in the three-signal BSS problem.

There are indications that this challenging path of further research could indeed prove fruitful. For instance, several aspects of the bicomplex-number algebra developed in Chapter 8 are strongly reminiscent of the quaternion algebra [1]. The resemblance between the bicomplex exponential (8.51) and the exponential form of quaternions [1, Section 12-7] is an example of this remarkable parallelism.

Appendix A

Higher-Order Statistics

The term **higher-order statistics (HOS)** refers to moments and cumulants of order greater than two. In many occasions, these higher-order moments and cumulants reveal more useful information than that extracted from only the first- and second-order statistics of the data (namely mean, variance and covariance). This appendix summarizes the fundamental facts about HOS. References [14,81,92,98–101] provide excellent reviews on HOS and their utility in signal processing applications. An extensive bibliography can be found in [121].

A.1 Definitions

A.1.1 Moments, cumulants and characteristic functions

Consider the set of random variables (rv's) $\{x_1, x_2, \ldots, x_n\}$. Its **first joint characteristic function** is defined as:

$$\phi(\omega_1, \omega_2, \ldots, \omega_n) \triangleq \mathrm{E}\big[\exp\big(j(\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n)\big)\big], \tag{A.1}$$

where j is the imaginary unit. Under certain conditions [120] ϕ can be expanded in power series about the origin:

$$\phi(\omega_1, \omega_2, \ldots, \omega_n) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} \mu_{i_1 i_2 \ldots i_n} \frac{(j\omega_1)^{i_1}\,(j\omega_2)^{i_2} \ldots (j\omega_n)^{i_n}}{i_1!\, i_2! \cdots i_n!}, \tag{A.2}$$

in which the coefficients $\mu_{i_1 i_2 \ldots i_n}$ are the **joint moments** of order $r = i_1 + i_2 + \cdots + i_n$ of $\{x_1, x_2, \ldots, x_n\}$:

$$\mu_{i_1 i_2 \ldots i_n} = \mathrm{Mom}_{i_1 i_2 \ldots i_n}[x_1, x_2, \ldots, x_n] \triangleq \mathrm{E}[x_1^{i_1} x_2^{i_2} \cdot \ldots x_n^{i_n}]. \tag{A.3}$$

From the expression (A.2) it can be deduced that:

$$\mu_{i_1 i_2 \ldots i_n} = (-j)^r \left. \frac{\partial^r \phi(\omega_1, \omega_2, \ldots, \omega_n)}{\partial \omega_1^{i_1} \, \partial \omega_2^{i_2} \ldots \partial \omega_n^{i_n}} \right|_{\omega_1 = \omega_2 = \cdots = \omega_n = 0} \tag{A.4}$$

Similarly, the **second joint characteristic function** of the set of rv's $\{x_1, x_2, \ldots, x_n\}$ is defined as the natural logarithm of its first characteristic function, that is:

$$\psi(\omega_1, \omega_2, \ldots, \omega_n) \triangleq \ln \phi(\omega_1, \omega_2, \ldots, \omega_n). \tag{A.5}$$

Its Taylor expansion about the origin, if it exists, yields:

$$\psi(\omega_1, \omega_2, \ldots, \omega_n) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} \kappa_{i_1 i_2 \ldots i_n} \frac{(j\omega_1)^{i_1} \, (j\omega_2)^{i_2} \ldots (j\omega_n)^{i_n}}{i_1! \, i_2! \ldots i_n!}, \tag{A.6}$$

where $\kappa_{i_1 i_2 \ldots i_n}$ is the **joint cumulant** of order $r = i_1 + i_2 + \cdots + i_n$:

$$\kappa_{i_1 i_2 \ldots i_n} = \mathrm{Cum}_{i_1 i_2 \ldots i_n}[x_1, x_2, \ldots, x_n], \tag{A.7}$$

which can also be calculated from ψ as:

$$\kappa_{i_1 \, i_2 \ldots i_n} = (-j)^r \left. \frac{\partial^r \psi(\omega_1, \omega_2, \ldots, \omega_n)}{\partial \omega_1^{i_1} \, \partial \omega_2^{i_2} \ldots \partial \omega_n^{i_n}} \right|_{\omega_1 = \omega_2 = \cdots = \omega_n = 0} \tag{A.8}$$

Strictly speaking, notation $\kappa^x_{i_1 \, i_2 \ldots i_n}$ should be used to indicate that the cumulant refers to rv's $\{x_1, x_2, \ldots, x_n\}$, and analogously for the moments $\mu^x_{i_1 \, i_2 \ldots i_n}$. However, for convenience the super-index will be suppressed in the sequel if no confusion is possible.

A.1.2 Notations

In occasions, an alternative but equivalent nomenclature for moments and cumulants may be more convenient. The relationship between these two notations can be expressed as:[1]

$$\mathrm{Cum}[\underbrace{x_{j_1}, \ldots, x_{j_1}}_{i_1}, \underbrace{x_{j_2}, \ldots, x_{j_2}}_{i_2}, \ldots, \underbrace{x_{j_n}, \ldots, x_{j_n}}_{i_n}] = \mathrm{Cum}_{i_1 i_2 \ldots i_n}[x_{j_1}, x_{j_2}, \ldots, x_{j_n}]. \tag{A.9}$$

[1] There exists a very interesting link between these variant statistical notations and those used in tensor algebra and the theory of homogeneous polynomials [51, 91].

For example, given the set of rv's $\{x_1, x_2, x_3\}$ and according to (A.9), we can write:

$$\mathrm{Cum}[x_1, x_2, x_3] = \mathrm{Cum}_{111}[x_1, x_2, x_3]$$
$$\mathrm{Cum}[x_1, x_1, x_2, x_3] = \mathrm{Cum}_{211}[x_1, x_2, x_3]$$
$$\mathrm{Cum}[x_1, x_2, x_2, x_3, x_3] = \mathrm{Cum}_{122}[x_1, x_2, x_3].$$

In this alternative nomenclature, $\kappa_{i_1 i_2 \ldots i_n}$ refers to $\mathrm{Cum}[x_{i_1}, x_{i_2}, \ldots, x_{i_n}]$, and not to $\mathrm{Cum}_{i_1 i_2 \ldots i_n}[x_1, x_2, \ldots, x_n]$ as it has been defined before in (A.7):

$$\kappa_{111} = \mathrm{Cum}[x_1, x_1, x_1] = \mathrm{Cum}_{300}[x_1, x_2, x_3] = \mathrm{Cum}_3[x_1]$$
$$\kappa_{122} = \mathrm{Cum}[x_1, x_2, x_2] = \mathrm{Cum}_{120}[x_1, x_2, x_3] = \mathrm{Cum}_{12}[x_1, x_2]$$
$$\kappa_{1223} = \mathrm{Cum}[x_1, x_2, x_2, x_3] = \mathrm{Cum}_{121}[x_1, x_2, x_3].$$

The notation being used depends on the context. For instance, when the symmetry and multilinearity properties of cumulants are defined in Appendix A.2, this alternative notation will be employed.

A.1.3 Relationship between moments and cumulants

All the cumulants of a set of rv's can be written as a function (usually non-linear) of the moments of the same set of rv's, and vice versa. Expressions for moments and cumulants and their existence conditions can be found in [120]. In general, we have [14,92]:

$$\mathrm{Cum}[x_1, x_2 \ldots, x_n] = \sum (-1)^{p-1}(p-1)! \prod_{i=1}^{p} \mathrm{E}\Big[\prod_{j\in\nu_i} x_j\Big], \tag{A.10}$$

where the summation extends over all possible p-group partitions $\{\nu_1, \nu_2, \ldots, \nu_p\}$ of the set of integers $\{1, 2, \ldots, n\}$, for $p = 1, \ldots, n$. According to this expression, the following are the formulae of the cumulants up to order four:

$$\mathrm{Cum}[x_1] = \mathrm{E}[x_1] \tag{A.11a}$$
$$\mathrm{Cum}[x_1, x_2] = \mathrm{E}[x_1 x_2] - \mathrm{E}[x_1]\mathrm{E}[x_2] \tag{A.11b}$$
$$\begin{aligned}\mathrm{Cum}[x_1, x_2, x_3] = {} & \mathrm{E}[x_1 x_2 x_3] - \big(\mathrm{E}[x_1]\mathrm{E}[x_2 x_3] + \mathrm{E}[x_2]\mathrm{E}[x_1 x_3] + \mathrm{E}[x_3]\mathrm{E}[x_1 x_2]\big) \\ & + 2\mathrm{E}[x_1]\mathrm{E}[x_2]\mathrm{E}[x_3]\end{aligned} \tag{A.11c}$$

$$\begin{aligned}\mathrm{Cum}[x_1,\,x_2,\,x_3,\,x_4] = {} & \mathrm{E}[x_1x_2x_3x_4] - \big(\mathrm{E}[x_1x_2]\mathrm{E}[x_3x_4] + \mathrm{E}[x_1x_3]\mathrm{E}[x_2x_4] \\ & + \mathrm{E}[x_1x_4]\mathrm{E}[x_2x_3] + \mathrm{E}[x_1]\mathrm{E}[x_2x_3x_4] + \mathrm{E}[x_2]\mathrm{E}[x_1x_3x_4] \\ & + \mathrm{E}[x_3]\mathrm{E}[x_1x_2x_4] + \mathrm{E}[x_4]\mathrm{E}[x_1x_2x_3]\big) + 2\big(\mathrm{E}[x_1]\mathrm{E}[x_2]\mathrm{E}[x_3x_4] \\ & + \mathrm{E}[x_1]\mathrm{E}[x_3]\mathrm{E}[x_2x_4] + \mathrm{E}[x_1]\mathrm{E}[x_4]\mathrm{E}[x_2x_3] + \mathrm{E}[x_2]\mathrm{E}[x_3]\mathrm{E}[x_1x_4] \\ & + \mathrm{E}[x_2]\mathrm{E}[x_4]\mathrm{E}[x_1x_3] + \mathrm{E}[x_3]\mathrm{E}[x_4]\mathrm{E}[x_1x_2]\big) \\ & - 6\mathrm{E}[x_1]\mathrm{E}[x_2]\mathrm{E}[x_3]\mathrm{E}[x_4]. \end{aligned} \tag{A.11d}$$

The **univariate** expressions ($n = 1$) are obtained by substituting $x_1 = \cdots = x_n = x$ above, giving:

$$\kappa_1 = \mathrm{Cum}[x] = \mu_1, \qquad \text{mean} \tag{A.12a}$$

$$\kappa_2 = \mathrm{Cum}[x,\,x] = \mu 2 - \mu_1^2, \qquad \text{variance} \tag{A.12b}$$

$$\kappa_3 = \mathrm{Cum}[x,\,x,\,x] = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3, \qquad \text{skewness} \tag{A.12c}$$

$$\kappa_4 = \mathrm{Cum}[x,\,x,\,x,\,x] = \mu_4 - 3\mu_2^2 - 4\mu_3\mu_1 + 12\mu_2\mu_1^2 - 6\mu_1^4, \qquad \text{kurtosis.} \tag{A.12d}$$

If the mean value is zero, the above expressions can be simplified into:

$$\kappa_1 = 0, \qquad \kappa_2 = \mu_2 \tag{A.13a}$$

$$\kappa_3 = \mu_3, \qquad \kappa_4 = \mu_4 - 3\mu_2^2. \tag{A.13b}$$

In the **bivariate** case, the cumulants of the rv's $\{x_1,\,x_2\}$ up to order four are given by [120]:

$$\text{1st order}: \quad \kappa_{10} = \mu_{10}, \quad \kappa_{01} = \mu_{01} \tag{A.14a}$$

$$\text{2nd order}: \quad \kappa_{20} = \mu_{20} - \mu_{10}^2, \quad \kappa_{11} = \mu_{11} - \mu_{10}\mu_{01}, \quad \kappa_{02} = \mu_{02} - \mu_{01}^2 \tag{A.14b}$$

$$\begin{aligned}\text{3rd order}: \quad & \kappa_{30} = \mu_{30} + 2\mu_{10}^3 - 3\mu_{20}\mu_{10} \\ & \kappa_{21} = \mu_{21} - \mu_{20}\mu_{01} - 2\mu_{11}\mu_{10} + 2\mu_{10}^2\mu_{01} \\ & \kappa_{12} = \mu_{12} - \mu_{02}\mu_{10} - 2\mu_{11}\mu_{01} + 2\mu_{01}^2\mu_{10} \\ & \kappa_{03} = \mu_{03} + 2\mu_{01}^3 - 3\mu_{02}\mu_{01} \end{aligned} \tag{A.14c}$$

$$
\begin{aligned}
\text{4th order}: \quad \kappa_{40} &= \mu_{40} - 3\mu_{20}^2 - 4\mu_{30}\mu_{10} - 6\mu_{10}^4 + 12\mu_{20}\mu_{10}^2 \\
\kappa_{31} &= \mu_{31} - \mu_{30}\mu_{01} - 3\mu_{21}\mu_{10} - 3\mu_{20}\mu_{11} + 6\mu_{11}\mu_{10}^2 + 6\mu_{20}\mu_{10}\mu_{01} \\
&\quad - 6\mu_{10}^3\mu_{01} \\
\kappa_{22} &= \mu_{22} - \mu_{20}\mu_{02} + 2(\mu_{20}\mu_{01}^2 + \mu_{02}\mu_{10}^2) - 2(\mu_{21}\mu_{01} + \mu_{12}\mu_{10} + \mu_{11}^2) - \\
&\quad - 6(\mu_{10}\mu_{01})^2 + 8\mu_{11}\mu_{10}\mu_{01} \\
\kappa_{13} &= \mu_{13} - \mu_{03}\mu_{10} - 3\mu_{12}\mu_{01} - 3\mu_{02}\mu_{11} + 6\mu_{11}\mu_{01}^2 + 6\mu_{02}\mu_{10}\mu_{01} \\
&\quad - 6\mu_{01}^3\mu_{10} \\
\kappa_{04} &= \mu_{04} - 3\mu_{02}^2 - 4\mu_{03}\mu_{01} - 6\mu_{01}^4 + 12\mu_{02}\mu_{01}^2. \qquad \text{(A.14d)}
\end{aligned}
$$

If $\mu_{10} = \mu_{01} = 0$ further simplifications produce [120]:

$$\text{1st order}: \quad \kappa_{10} = 0, \quad \kappa_{01} = 0 \qquad \text{(A.15a)}$$

$$\text{2nd order}: \quad \kappa_{20} = \mu_{20}, \quad \kappa_{11} = \mu_{11}, \quad \kappa_{02} = \mu_{02} \qquad \text{(A.15b)}$$

$$
\begin{aligned}
\text{3rd order}: \quad & \kappa_{30} = \mu_{30}, \quad \kappa_{21} = \mu_{21} \\
& \kappa_{12} = \mu_{12}, \quad \kappa_{03} = \mu_{03} \qquad \text{(A.15c)}
\end{aligned}
$$

$$
\begin{aligned}
\text{4th order}: \quad & \kappa_{40} = \mu_{40} - 3\mu_{20}^2, \quad \kappa_{31} = \mu_{31} - 3\mu_{20}\mu_{11} \\
& \kappa_{22} = \mu_{22} - \mu_{20}\mu_{02} - 2\mu_{11}^2 \\
& \kappa_{13} = \mu_{13} - 3\mu_{02}\mu_{11}, \quad \kappa_{04} = \mu_{04} - 3\mu_{02}^2. \qquad \text{(A.15d)}
\end{aligned}
$$

Therefore, in such a case the cumulants up to order 3 are equal to the corresponding moments. κ_{mn} with $m = 0$ or $n = 0$ is called a **marginal cumulant**, since it only involves one r.v., while if $m \neq 0$, $n \neq 0$, is an $(m+n)$th-order **cross-cumulant** of the rv's $\{x_1, x_2\}$. Analogue expressions and nomenclature hold in the case of more than two rv's [101,120].

A.1.4 Skewness and kurtosis. Tailness

The third-order marginal cumulant (A.12c) is also known as **skewness**, and characterizes the degree of asymmetry of a distribution around its mean. Hence, this statistic is zero for variables with symmetrical probability density functions (pdf's). The fourth-order marginal cumulant (A.12d) is also called **kurtosis**. It measures the peakedness or flatness — also known as the **tailness** — of a distribution in relation to the Gaussian pdf. A distribution with positive kurtosis is named *leptokurtic, long-tailed* or *super-Gaussian* [114] (e.g., exponential). A distribution with negative kurtosis is called *platykurtic, short-tailed* or *sub-Gaussian* [114] (e.g., uniform). A zero-kurtosis distribution is referred to as *mesokurtic* (e.g., Gaussian). These concepts [120] are illustrated in Figure A.1.

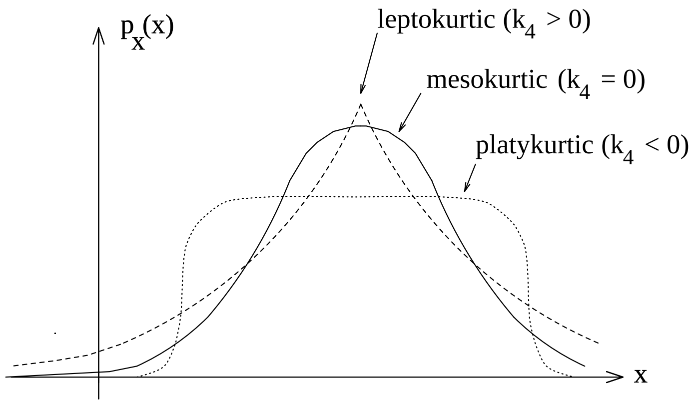


Figure A.1: Types of distribution as a function of the kurtosis sign or pdf tail.

A.2 Properties

The most important properties of moments and cumulants are summarized below [14, 91, 92, 98, 100, 120]. Although the results are presented for real-valued variables, the extension to the complex case is straightforward through the appropriate inclusion of the complex-conjugate operator. Given the set of rv's $\{x_1, x_2, \ldots, x_n\}$, $x_i \in \mathbb{R}$, $i = 1, \ldots, n$:

1. Scaling:

$$\mathrm{Mom}_{i_1 i_2 \ldots i_n}[a_1 x_1,\, a_2 x_2, \ldots,\, a_n x_n] = \left(\prod_{k=1}^{n} a_k^{i_k}\right) \mathrm{Mom}_{i_1 i_2 \ldots i_n}[x_1,\, x_2, \ldots,\, x_n]$$

$$\mathrm{Cum}_{i_1 i_2 \ldots i_n}[a_1 x_1,\, a_2 x_2, \ldots,\, a_n x_n] = \left(\prod_{k=1}^{n} a_k^{i_k}\right) \mathrm{Cum}_{i_1 i_2 \ldots i_n}[x_1,\, x_2, \ldots,\, x_n]$$

$$\forall (i_1,\, i_2, \ldots,\, i_n) \in \mathbb{N}^n, \quad \forall (a_1,\, a_2, \ldots,\, a_n) \in \mathbb{R}^n.$$

2. Symmetry: moments and cumulants are symmetric functions in their arguments, e.g.:

$$\mathrm{Mom}[x_1,\, x_2,\, x_3] = \mathrm{Mom}[x_1,\, x_3,\, x_2] = \mathrm{Mom}[x_2,\, x_1,\, x_3] = \dots$$
$$\mathrm{Cum}[x_1,\, x_2,\, x_3] = \mathrm{Cum}[x_1,\, x_3,\, x_2] = \mathrm{Cum}[x_2,\, x_1,\, x_3] = \dots$$

3. Invariability: a change in the mean value leaves the cumulants unchanged, with the sole exception of the first-order cumulant, which is equal to the mean. Hence:

$$\mathrm{Cum}_{i_1 i_2 \dots i_n}[x_1 + a_1,\, x_2 + a_2, \dots,\, x_n + a_n] = \mathrm{Cum}_{i_1 i_2 \dots i_n}[x_1,\, x_2, \dots,\, x_n]$$
$$\forall (a_1,\, a_2, \dots,\, a_n) \in \mathbb{R}^n, \quad \forall (i_1,\, i_2, \dots,\, i_n) \in \mathbb{N}^n \,/\, i_1 + i_2 + \dots + i_n \neq 1,$$

whereas

$$\mathrm{Mom}_{i_1 i_2 \dots i_n}[x_1 + a_1,\, x_2 + a_2, \dots,\, x_n + a_n] \neq \mathrm{Mom}_{i_1 i_2 \dots i_n}[x_1,\, x_2, \dots,\, x_n],$$

in general.

4. Independence: if the sets of rv's $\{x_1,\, x_2, \dots,\, x_n\}$ and $\{y_1,\, y_2, \dots,\, y_n\}$ are independent, then, $\forall (i_1,\, i_2, \dots,\, i_n) \in \mathbb{N}^n$:

$$\mathrm{Cum}_{i_1 i_2 \dots i_n}[x_1 + y_1,\, x_2 + y_2, \dots,\, x_n + y_n] =$$
$$= \mathrm{Cum}_{i_1 i_2 \dots i_n}[x_1,\, x_2, \dots,\, x_n] + \mathrm{Cum}_{i_1 i_2 \dots i_n}[y_1,\, y_2, \dots,\, y_n],$$

whereas, in general:

$$\mathrm{Mom}_{i_1 i_2 \dots i_n}[x_1 + y_1,\, x_2 + y_2, \dots,\, x_n + y_n] \neq$$
$$\neq \mathrm{Mom}_{i_1 i_2 \dots i_n}[x_1,\, x_2, \dots,\, x_n] + \mathrm{Mom}_{i_1 i_2 \dots i_n}[y_1,\, y_2, \dots,\, y_n].$$

5. Gaussianity: if the set of rv's $\{x_1,\, x_2, \dots,\, x_n\}$ is jointly Gaussian, then all their cumulants of order $r > 2$ are null:

$$\mathrm{Cum}_{i_1 i_2 \dots i_n}[x_1,\, x_2, \dots,\, x_n] = 0, \qquad \forall (i_1,\, i_2, \dots,\, i_n) \in \mathbb{N}^n \,/\, i_1 + i_2 + \dots + i_n > 2.$$

6. Multilinearity: if the random vectors $\mathbf{x} = [x_1, x_2, \ldots, x_q]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, y_2, \ldots, y_p]^{\mathrm{T}}$ are related through a linear transformation $W = (w_{ij}) \in \mathbb{R}^{p \times q}$ so that $\mathbf{y} = W\mathbf{x}$, then, $\forall r \in \mathbb{N}^+$:

$$\mathrm{Cum}[y_{i_1}, y_{i_2}, \ldots, y_{i_r}] = \sum_{j_1=1}^{q} \sum_{j_2=1}^{q} \cdots \sum_{j_r=1}^{q} w_{i_1 j_1} w_{i_2 j_2} \ldots w_{i_r j_r} \, \mathrm{Cum}[x_{j_1}, x_{j_2}, \ldots, x_{j_r}].$$

By virtue of Properties 1 and 3, the cumulants are called *semi-invariants*, since they are not affected by a change in the mean value (except at first order). From Property 5, it follows that all the statistical properties of Gaussian rv's are defined entirely by their first- and second-order cumulants. Therefore, the cumulants of order greater than two, in some sense, measure the non-Gaussian nature of a rv. Finally, thanks to Property 6, cumulants can be regarded as *tensors* [91]. This means that the rth-order cumulants of a random vector linearly related to another random vector are also linearly related to the rth-order cumulants of the latter. The coefficients of this relationship are non-linear combinations of the elements of the transformation linking both vectors.

A.3 Statistical Independence Criterion Based on Cumulants

Theorem (cumulant-based independence criterion). *The random variables* $\{x_1, x_2, \ldots, x_n\}$ *are statistically independent if and only if all their cross-cumulants are equal to zero:*

$$\mathrm{Cum}_{i_1 i_2 \ldots i_n}[x_1, x_2, \ldots, x_n] = 0, \qquad \forall (i_1, i_2, \ldots, i_n) \in \mathbb{P}^n, \tag{A.16a}$$

where

$$\mathbb{P}^n = \big\{ (i_1, i_2, \ldots, i_n) \in \mathbb{N}^n \ : \nexists !\, i_k \neq 0,\ k = 1, 2, \ldots, n \big\}. \tag{A.16b}$$

Proof. The set of rv's $\{x_1, x_2, \ldots, x_n\}$ are statistically independent if and only if its joint pdf (jpdf) can be expressed as the product of the marginals pdf's. In terms of the characteristic functions, this is equivalent to:

$$\phi(\omega_1, \omega_2, \ldots, \omega_n) = \phi_1(\omega_1)\phi_2(\omega_2)\cdots\phi_n(\omega_n),$$
$$\phi_i(\omega_i) = \mathrm{E}[\exp(j\omega_i x_i)], \qquad i = 1, 2, \ldots, n,$$

or, taking logarithms, to:

$$\psi(\omega_1, \omega_2, \ldots, \omega_n) = \psi_1(\omega_1) + \psi_2(\omega_2) + \cdots + \psi_n(\omega_n). \tag{A.17}$$

Now, obtaining the Taylor series expansion about the origin of both sides of (A.17), we have:

$$\sum_{i_1}\sum_{i_2}\cdots\sum_{i_n} \kappa_{i_1 i_2 \ldots i_n} \frac{(j\omega_1)^{i_1}\,(j\omega_2)^{i_2}\ldots(j\omega_n)^{i_n}}{i_1!\,i_2!\ldots i_n!} =$$
$$= \sum_{p_1} \kappa_{p_1} \frac{(j\omega_1)^{p_1}}{p_1!} + \sum_{p_2} \kappa_{p_2} \frac{(j\omega_2)^{p_2}}{p_2!} + \cdots + \sum_{p_n} \kappa_{p_n} \frac{(j\omega_n)^{p_n}}{p_n!}. \tag{A.18}$$

On the right-hand side of (A.18) only marginal cumulants appear, so this expression is true for all $(\omega_1, \omega_2, \ldots, \omega_n) \in \mathbb{R}^n$ if and only if all the cross-cumulants are null. ■

Generalizing this result, a set of rv's are said to be *statistically independent up to order* r if and only if all their sth-order cross-cumulants are null, for $s \leqslant r$. In particular, uncorrelatedness is tantamount to second-order independence. For Gaussian variables, hence, uncorrelatedness and independence are equivalent.

A.4 Standardized Moments and Cumulants of Several Distributions

Tables A.1 and A.2 show the values of the moments and cumulants, respectively, for several well-known distributions.[2] It is assumed that the mean is zero and the variance one, hence the term *standardized* or *normalized*. The values for a variable of arbitrary mean μ and variance σ^2 can be calculated via:

$$\mu_n = \sum_{r=0}^{n} \binom{n}{r} \sigma^r \tilde{\mu}_r \mu^{n-r}, \tag{A.19}$$

[2]The notation "Triangular (a, b)" represents a triangular pdf $p(x)$ defined between $x = a$ and $x = b$ with maximum value at $x = 0$.

in which $\tilde{\mu}_0 = 1$. As to the cumulants:

$$\kappa_n = \sigma^n \tilde{\kappa}_n, \tag{A.20}$$

with $\kappa_1 = \mu$.

MOMENT VALUES FOR SEVERAL DISTRIBUTIONS

DISTRIBUTION	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\mu}_3$	$\tilde{\mu}_4$	$\tilde{\mu}_5$	$\tilde{\mu}_6$	$\tilde{\mu}_7$	$\tilde{\mu}_8$
Exponential	0	1	2	9	44	265	1854	14833
Laplacian	$\tilde{\mu}_n = 0, \;\; n$ odd; $\quad \tilde{\mu}_n = n!/2^{n/2}, \;\; n$ even							
Rayleigh	0	1	0.6311	3.2451	5.9980	21.7911	64.8166	238.8266
Gaussian	0	1	0	3	0	15	0	105
Triangular $(-1,\, 7)^2$	0	1	0.5324	12/5	3.0421	8.7265	15.9713	40.1369
Uniform	0	1	0	9/5	0	27/7	0	9
Sinusoid	0	1	0	3/2	0	5/2	0	35/8
PRBS (B.1), $p = 0.5$	$\tilde{\mu}_n = 0, \;\; n$ odd; $\quad \tilde{\mu}_n = 1, \;\; n$ even							
PRBS (B.1), any p	$\tilde{\mu}_n = [-\sqrt{p/(1-p)}]^n (1-p) + [\sqrt{(1-p)/p}]^n p$							

Table A.1: Normalized moments of some distributions.

CUMULANT VALUES FOR SEVERAL DISTRIBUTIONS

DISTRIBUTION	$\tilde{\kappa}_1$	$\tilde{\kappa}_2$	$\tilde{\kappa}_3$	$\tilde{\kappa}_4$	$\tilde{\kappa}_5$	$\tilde{\kappa}_6$	$\tilde{\kappa}_7$	$\tilde{\kappa}_8$
Exponential	0	$\tilde{\kappa}_n = (n-1)!, \quad n > 1$						
Laplacian	0	1	0	3	0	30	0	630
Rayleigh	0	1	0.6311	0.2451	-0.3131	-0.8683	-0.2879	4.1103
Gaussian	0	1	$\tilde{\kappa}_n = 0, \quad n > 2$					
Triangular $(-1,\, 7)^2$	0	1	0.5324	$-3/5$	-2.2816	-0.1077	19.1655	40.2162
Uniform	0	1	0	$-6/5$	0	$48/7$	0	$-432/5$
Sinusoid	0	1	0	$-3/2$	0	10	0	$-1155/8$
PRBS (B.1), $p = 0.5$	0	1	0	-2	0	16	0	-272
PRBS (B.1), any p	0	1	See Table B.3.					

Table A.2: Normalized cumulants of some distributions.

Appendix B

Modelling Signals of Arbitrary Kurtosis via Binary Sequences

B.1 Introduction

A myriad of methods for BSS (such as [29, 41, 48, 69, 88], to cite just a few references; see Chapter 2 for a detailed account) rely on the HOS of the available data, and in particular on their fourth-order cumulants. As a consequence, the performance of these separation techniques strongly depends on the sign or value of the source (normalized) kurtosis. Aiming at carrying out a thorough testing of these procedures, it is thus desirable to have available distributions with a range of kurtosis values.

Consider the Bernoulli rv, characterized by the events $x = a$ and $x = b$ ($b > a$) of probability:

$$P(x) = \begin{cases} 1 - p, & x = a \\ p, & x = b. \end{cases} \tag{B.1}$$

In this appendix (which is an updated, expanded version of [135]) it is proved that such distribution presents a normalized kurtosis lying in the range $[-2, +\infty[$ as the parameter p varies along the probability interval $[0, 1]$; that is, this distribution is able to produce all allowed values for the statistic of interest. This outcome suggests that straightforward binary sequences (also called *pseudorandom binary sequences, PRBS's*)

DISTRIBUTION	$\tilde{\kappa}_4$
Uniform-phase sinusoid	-1.5
Uniform	-1.2
Gaussian	0
Rayleigh	0.2451
Laplacian	3
Exponential	6

Table B.1: Normalized kurtosis values of various random distributions.

of suitable probabilities may be used instead of other more complicated source distributions to assess the performance of fourth-order cumulant based BSS methods. In the complex case, 4-QAM signals [108] of appropriate probability distribution fulfil analogous objectives.

B.2 Real Case: The Bernoulli Distribution

The moments and cumulants up to the fourth order of the Bernoulli rv (B.1) are summarized in Table B.2. The normalized kurtosis is defined as

$$\tilde{\kappa}_4 = \frac{\text{kurtosis}}{\text{variance}^2} = \frac{\kappa_4}{\kappa_2^2}. \tag{B.2}$$

This is equivalent to the evaluation of the kurtosis when the variance of the distribution is unity, i.e., when the rv has been normalized to unit variance. Table B.1 shows the value of this parameter for several well-known random distributions (extracted from Table A.2). Some normalized cumulants of the Bernoulli distribution appear in Table B.3. In particular, at 4th-order we have:

$$\tilde{\kappa}_4 = \frac{1 - 6p + 6p^2}{p\,(1-p)} = \frac{6(p - p_1)(p - p_2)}{p\,(1-p)}, \tag{B.3}$$

p_1 and p_2 being the roots of the polynomial in the numerator of $\tilde{\kappa}_4$, given by

$$p_1,\, p_2 = \frac{1}{2}\left(1 \pm \frac{1}{\sqrt{3}}\right). \tag{B.4}$$

Figure B.1 plots $\tilde{\kappa}_4$ against p. In that plot it is seen how the normalized kurtosis takes all values from -2 to $+\infty$ as p varies across its allowed range [0, 1]. Furthermore,

MOMENTS	CUMULANTS
$\mu_1 = a\,(1-p) + b\,p$	$\kappa_1 = \mu_1$
$\mu_2 = a^2(1-p) + b^2 p$	$\kappa_2 = (b-a)^2 p\,(1-p)$
$\mu_3 = a^3(1-p) + b^3 p$	$\kappa_3 = (b-a)^3 p\,(1-p)(1-2p)$
$\mu_4 = a^4(1-p) + b^4 p$	$\kappa_4 = (b-a)^4 p\,(1-p)(1-6p+6p^2)$

Table B.2: Statistics up to the 4th-order of the Bernoulli distribution.

CUMULANT	VALUE
$\tilde{\kappa}_3$	$\frac{1-2p}{\sqrt{p(1-p)}}$
$\tilde{\kappa}_4$	$\frac{1-6p+6p^2}{p(1-p)}$
$\tilde{\kappa}_5$	$\frac{(1-2p)(1-12p+12p^2)}{\sqrt{p(1-p)}^3}$
$\tilde{\kappa}_6$	$\frac{1-30p+150p^2-240p^3+120p^4}{[p(1-p)]^2}$
$\tilde{\kappa}_7$	$\frac{(1-2p)(1-60p+420p^2-720p^3+360p^4)}{\sqrt{p(1-p)}^5}$
$\tilde{\kappa}_8$	$\frac{1-126p+1806p^2-8400p^3+16800p^4-15120p^5+5040p^6}{[p(1-p)]^3}$

Table B.3: Normalized cumulants ($\tilde{\kappa}_1 = 0$, $\tilde{\kappa}_2 = 1$) of the Bernoulli (PRBS) distribution (B.1).

it is easy to show [120, page 107] that $\tilde{\kappa}_4 \geqslant -2$ for any distribution. The equality holds true for the only case of a Bernoulli pdf with $p = 0.5$.

Therefore, with an appropriate choice of the parameter p, the Bernoulli rv is able to provide any normalized kurtosis value. Specifically, from equations (B.3) and (B.4), the value of probability p necessary to obtain a desired $\tilde{\kappa}_4$ is given by:

$$p = \frac{1}{2}\left(1 \pm \sqrt{\frac{\tilde{\kappa}_4 + 2}{\tilde{\kappa}_4 + 6}}\right). \tag{B.5}$$

A Bernoulli distribution naturally defines a binary stochastic process. Hence, the above results implies that it is possible to model *any* signal distribution, as far as $\tilde{\kappa}_4$ is concerned, by means of a simple binary sequence (PRBS) of proper probability. From

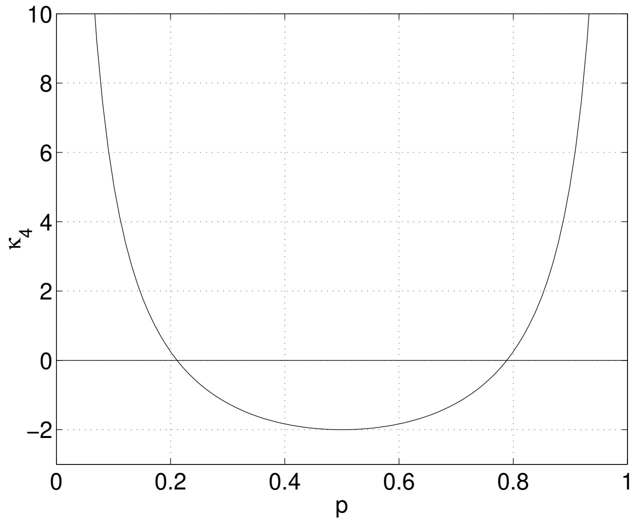


Figure B.1: Normalized kurtosis versus event probability in the Bernoulli random variable.

the BSS standpoint, this observation seems to indicate that PRBS's may be employed in the place of continuous-distribution source random processes in order to evaluate BSS methods. This reasonable hypothesis is empirically checked, and indeed validated, in Appendix B.4. But first, we extend these results to the complex case.

B.3 Complex Signals

Consider the complex rv $z = x + jy \in \mathbb{C}$ with x and y independent real rv's, and j being the imaginary unit. Due to the multilinearity property of cumulants (Property 6 in Appendix A.2), $\kappa_n^z = \kappa_n^x + \kappa_n^y$, $\forall n \in \mathbb{N}^+$, n even (here the cumulants are defined with the same number of conjugated and non-conjugated variables). Now, assume that x and y have identical distribution, and hence, identical cumulants at any order. In these conditions, the normalized kurtosis of z is then given by:

$$\tilde{\kappa}_4^z \triangleq \frac{\kappa_4^z}{(\kappa_2^z)^2} = \frac{\kappa_4^x + \kappa_4^y}{(\kappa_2^x + \kappa_2^y)^2} = \frac{\kappa_4^x}{2(\kappa_2^x)^2} = \frac{\tilde{\kappa}_4^x}{2}, \tag{B.6}$$

i.e., it is half the normalized kurtosis of each of its real components. Therefore, if these components are independent Bernoulli variables characterized by identical probability p, the normalized kurtosis of the resulting complex variable can range from -1 to ∞

when p is varied accordingly. Remark that the resulting complex processes constitute particular 4-QAM signals.

B.4 Experimental Validation

The above claims are illustrated and validated on an adaptive source separation algorithm: Cardoso-Laheld's EASI method [29]. Essentially, EASI employs the following adaptation rule for the separating matrix:

$$W_{k+1} = \big(I - \lambda_k H_{\mathbf{g}}(\mathbf{s}_k)\big)\, W_k. \tag{B.7}$$

This rule reduces to the specification of a set of non-linear functions $\mathbf{g}(\cdot)$ of the estimated sources $\mathbf{s}_k = W_k \mathbf{y}_k$. There exists a very close relationship among these functions, the sources statistics and the stability of the algorithm. In fact, for the case of cubic nonlinearities, the convergence condition becomes that the sum of normalized kurtosis of any two sources must be less than zero:

$$\kappa_4^{x_i} + \kappa_4^{x_j} < 0, \qquad 1 \leqslant i < j \leqslant q, \tag{B.8}$$

$\kappa_4^{x_i}$ being the ith-source normalized kurtosis.

For the following simulations, a two-source two-sensor scenario is chosen, with a fixed invertible mixing matrix $M = \left[\begin{smallmatrix} 0.2 & -0.3 \\ 0.1 & 0.4 \end{smallmatrix}\right]$. A fixed value of the adaptation step is also selected as $\lambda_k = \lambda = 0.01$. The choice of cubic nonlinearities and uniformly distributed independent sources meets the stability condition (B.8). The identity matrix is taken as the initial separating matrix. The trajectories of the square global matrix elements as a function of the iteration number (equal to the sample number, in this case) are recorded for each separation. Separations are obtained in 100 Monte Carlo runs with these conditions, but with different (independent) source realizations at each run. Finally, the trajectories are averaged over the 100 trials. Figure B.2 displays the results obtained for the first 500 iterations. The global matrix diagonal elements converge to 1, while the two off-diagonal elements converge to 0, as expected for a successful source extraction.

Next, the uniformly distributed sources are substituted by two independent binary sequences with the same normalized kurtosis value as the uniform distribution, $\tilde{\kappa}_4 =$

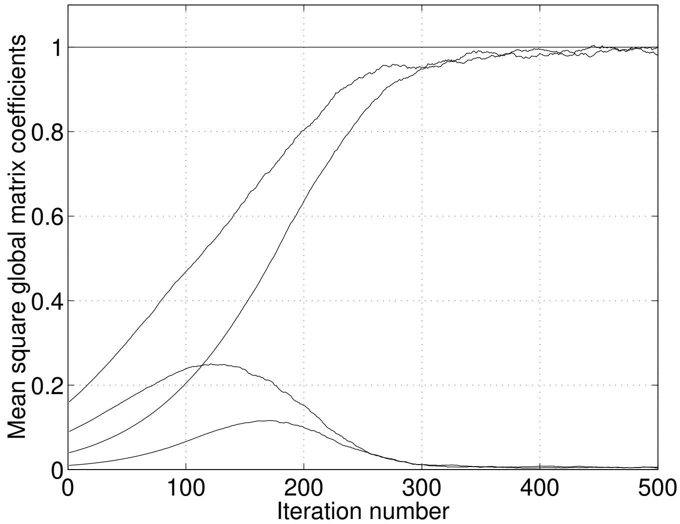


Figure B.2: Trajectories averaged over 100 Monte Carlo runs of the square global matrix elements obtained by the EASI algorithm for two uniformly distributed sources. Cubic nonlinearities, $\lambda_k = 0.01$.

-1.2 (Table B.1). From expression (B.5), the probability of one of their two levels is obtained as $p = 0.7041$.[1] The above experiment is carried out once again, in the same conditions but with the new set of sources and over a new independent set of 100 Monte Carlo trials. The trajectories of the mean square value of the global system elements are represented in Figure B.3, again for the first 500 iterations.

B.5 Discussion

Although the actual waveforms and pdf shapes inputted to the algorithm are totally different, Figs. B.2 and B.3 are essentially identical. This and further experiments thus confirm that for performance testing specific distributions may be simulated by means of PRBS's and an appropriate probability choice. On the other hand, a continuous pdf with a desired value of $\tilde{\kappa}_4$ may be difficult to find or may indeed not exist. By contrast, according to expression (B.5), there always exists a value of p that provides the required statistic value. The potential of the proposed technique is further endorsed by the fact

[1] Note that the sign choice in eqn. (B.5) is immaterial, since the two resulting probability values p_1 and p_2 verify $p_2 = 1 - p_1$.

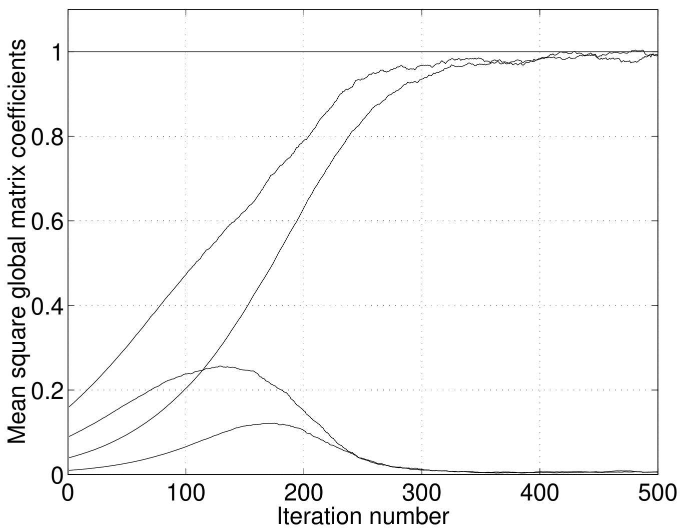


Figure B.3: Trajectories averaged over 100 Monte Carlo trials of the square global matrix coefficients obtained by the EASI method for two PRBS sources with the same normalized kurtosis as a uniform distribution ($p = 0.7041$). Cubic nonlinearities, $\lambda_k = 0.01$.

that the statistic of interest can be varied across its allowed range in a gradual fashion.

Other techniques to obtain a range of kurtosis values are reported in [95, Section 3.4] and [127]. In the former, three-symbol symmetric, hence of fixed null skewness, discrete distributions are used. In the latter, symmetric pdf's are simulated via the product of uniform rv's. Our approach can prove more versatile in certain applications, as it also allows the simulation of other distribution parameters (Table B.3). In addition, signals are easier to generate, by simple thresholding of the output of a uniform random-number generator.

As a final remark, it should be kept in mind that this procedure for simulating distributions via binary sequences to assess blind separation techniques is only meaningful as long as the BSS method under testing does not take advantage of the discrete nature of the signals at its input.

Bibliography

[1] S. L. Altmann, *Rotations, Quaternions and Double Groups*, Oxford: Oxford University Press, 1986.

[2] S. Amari and A. Cichocki, "Adaptive Blind Signal Processing — Neural Network Approaches", *Proceedings of the IEEE*, Vol. 86, No. 10, October 1996, pp. 2026–2048.

[3] S. Amari, "Superefficiency in Blind Source Separation", *IEEE Transactions on Signal Processing*, Vol. 47, No. 4, April 1999, pp. 936–944.

[4] P.-O. Amblard and J.-M. Brossier, "Adaptive Estimation of the Fourth-Order Cumulant of a White Stochastic Process", *Signal Processing*, Vol. 42, 1995, pp. 37–43.

[5] P.-O. Amblard, M. Gaeta and J.-L. Lacoume, "Statistics for Complex Variables and Signals — Part I: Variables", *Signal Processing*, Vol. 53, 1996, pp. 1–13.

[6] E. Bacharakis, *Separation of Maternal and Foetal ECG Using Blind Source Separation Methods*, MSc Thesis, University of Strathclyde, Glasgow, Scotland, UK, September 1995.

[7] E. Bacharakis, A. K. Nandi and V. Zarzoso, "Foetal ECG Extraction Using Blind Source Separation Methods", in: *Proceedings EUSIPCO-96, VIII European Signal Processing Conference*, Trieste, Italy, September 10–13, 1996, Vol. I, pp. 395–398.

[8] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind

Separation and Blind Deconvolution", *Neural Computation*, Vol. 7, No. 6, 1995, pp. 1129–1159.

[9] A. Belouchrani and J.-F. Cardoso, "Maximum Likelihood Source Separation for Discrete Sources", in: M. Holt *et al.* (Eds.), *Proceedings EUSIPCO-94, VII European Signal Processing Conference*, Edinburgh, Scotland, UK, September 1994, pp. 768–771.

[10] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics", *IEEE Transactions on Signal Processing*, Vol. 45, No. 2, February 1997, pp. 434–444.

[11] A. Benveniste, M. Métivier, Michel et P. Priouret, *Algorithmes Adaptatifs et Approximations Stochastiques*, Paris: Masson, 1987 (English translation by S. S. Wilson, *Adaptive Algorithms and Stochastic Approximations*, Berlin: Springer-Verlag, 1990.)

[12] E. Biglieri and K. Yao, "Some Properties of Singular Value Decomposition and their Applications to Digital Signal Processing", *Signal Processing*, Vol. 18, No. 3, November 1989, pp. 277–289.

[13] R. E. Bogner, *Blind Separation of Sources*, Memo. No. 4559, Defence Evaluation & Research Agency, Malvern, UK, May 1992.

[14] D. R. Brillinger, *Time Series. Data Analysis and Theory*, San Francisco: Holden-Day, 1981.

[15] D. Callaerts, B. De Moor, J. Vandewalle, W. Sansen, G. Vantrappen and J. Janssens, "Comparison of SVD Methods to Extract the Foetal Electrocardiogram from Cutaneous Electrode Signals", *Medical & Biological Engineering & Computing*, May 1990, Vol. 28, pp. 217–224.

[16] X.-R. Cao and R. Liu, "General Approach to Blind Source Separation", *IEEE Transactions on Signal Processing*, Vol. 44, No. 3, March 1996, pp. 562–571.

[17] J.-F. Cardoso, "Source Separation Using Higher-Order Moments", in: *Proceedings ICASSP-89, IEEE International Conference on Acoustics, Speech and Signal Processing*, Glasgow, Scotland, UK, May 23–26, 1989, pp. 2109–2112.

[18] J.-F. Cardoso,"Blind Identification of Independent Components with Higher-Order Statistics", in: *Proceedings of the Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, USA, June 28–30, 1989, pp. 157–160.

[19] J.-F. Cardoso and P. Comon, "Tensor-Based Independent Component Analysis", in: L. Torres *et al.* (Eds.), *Proceedings EUSIPCO-90, V European Signal Processing Conference*, Barcelona, Spain, 1990, pp. 673–676.

[20] J.-F. Cardoso, "Eigen-Structure of the Fourth-Order Cumulant Tensor with Application to the Blind Source Separation Problem", *Proceedings ICASSP-90, IEEE International Conference on Acoustics, Speech and Signal Processing*, Albuquerque, New Mexico, USA, April 3–6, 1990, Vol. 5, pp. 2655–2658.

[21] J.-F. Cardoso, "Super-Symmetric Decomposition of the Fourth-Order Cumulant Tensor. Blind Identification of More Sources than Tensors", in: *Proceedings ICASSP-91, IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, May 14–17, 1991, Vol. 5, pp. 3109–3112.

[22] J.-F Cardoso, "Iterative Techniques for Blind Source Separation Using Only Fourth-Order Cumulants", in: J. Vandewalle *et al.* (Eds.), *Proceedings EUSIPCO-92, VI European Signal Processing Conference*, Brussels, 1992, pp. 739–742.

[23] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming for non-Gaussian Signals", *IEE Proceedings-F*, Vol. 140, No. 6, December 1993, pp. 362–370.

[24] J.-F. Cardoso, "On the Performance of Orthogonal Source Separation Algorithms", in: M. Holt *et al.* (Eds.), *Proceedings EUSIPCO-94, VII European Signal Processing Conference*, Edinburgh, Scotland, UK, September 1994, pp. 776–779.

[25] J.-F. Cardoso, "The Invariant Approach to Source Separation", in: *Proceedings NOLTA'95*, 1995, pp. 55–60.

[26] J.-F. Cardoso, S. Bose and B. Friedlander, "Output Cumulant Matching for Source Separation", in: *Proceedings IEEE Signal Processing Workshop on Higher-Order Statistics*, Aiguablava, Spain, 1995, pp. 44–48.

[27] J.-F. Cardoso, "Performance and Implementation of Invariant Source Separation Algorithms", in: *Proceedings ISCAS'96*, 1996.

[28] J.-F. Cardoso, S. Bose and B. Friedlander, "On Optimal Source Separation Based on Second and Fourth Order Cumulants", in: *Proceedings of the IEEE Statistical Signal and Array Processing Workshop*, Corfu, Greece, 1996.

[29] J.-F. Cardoso and B. H. Laheld, "Equivariant Adaptive Source Separation", *IEEE Transactions on Signal Processing*, Vol. 44, No. 12, December 1996, pp. 3017–3030.

[30] J.-F. Cardoso, "Estimating Equations for Source Separation", in: *Proceedings ICASSP-97, IEEE International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, April 21–24, 1997, Vol. V, pp. 3449–3452.

[31] J.-F. Cardoso, "Infomax and Maximum Likelihood in Blind Source Separation", *IEEE Signal Processing Letters*, Vol. 4, No. 4, April 1997, pp. 112–114.

[32] J.-F. Cardoso, "Statistical Principles of Source Separation", in: *Proceedings IFAC SYSID'97*, Fukuoka, Japan, pp. 1837–1844.

[33] J.-F. Cardoso, "Multidimensional Independent Component Analysis", in: *Proceedings ICASSP-98, IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, May 1998, Vol. IV, pp. 1941–1944.

[34] J.-F. Cardoso, "Blind Signal Separation: Statistical Principles", *Proceedings of the IEEE*, Vol. 86, No. 10, October 1998, pp. 2009–2025.

[35] J.-F. Cardoso, C. Jutten and P. Loubaton (Eds.), *Proceedings ICA'99, First International Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, January 11–15, 1999.

[36] B. W. Char, K. O. Geddes, G. H. Gonnet, B. L. Leong, M. B. Monagan and S. M. Watt, *Maple V Library Reference Manual*, New York: Springer-Verlag, 1991.

[37] E. Chaumette, P. Comon and D. Muller, "ICA-Based Technique for Radiating Sources Estimation: Application to Airport Surveillance", *IEE Proceedings-F*, Vol. 140, No. 6, December 1993, pp. 395–401.

[38] E. C. Cherry, "Some Experiments on the Recognition of Speech with One and Two Ears", *Journal of the Acoustical Society of America*, Vol. 25, 1953, pp. 975–979.

[39] I. J. Clarke, "Direct Exploitation of non-Gaussianity as a Discriminant", *Proceedings EUSIPCO-98, IX European Signal Processing Conference*, Rhodes, Greece, September 8–11, 1998, Vol. IV, pp. 2057–2060.

[40] R. M. Clemente and J. I. Acha, "Blind Separation of Sources Using a New Polynomial Equation", *IEE Electronics Letters*, Vol. 33, No. 3, 30th January 1997, pp. 176–177.

[41] P. Comon, "Separation of Stochastic Processes", in: *Proceedings of the Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, USA, June 28–30, 1989, pp. 174–179.

[42] P. Comon, "Separation of Sources Using Higher-Order Cumulants", *SPIE Vol. 1152 Advanced Algorithms and Architectures for Signal Processing IV*, 1989, pp. 170–181.

[43] P. Comon, "Statistical Approach to the Jutten-Herault Algorithm", in: *Proceedings of the NATO Workshop on Neuro-Computing*, Les Arcs, France, February 1989.

[44] P. Comon, "Higher-Order Separation, Application to Detection and Localization", in: L. Torres *et al.* (Eds.), *Proceedings EUSIPCO-90, V European Signal Processing Conference*, Barcelona, Spain, 1990, pp. 277–280.

[45] P. Comon, "Analyse en Composantes Indépendantes et Identification Aveugle", *Traitement du Signal*, Vol. 7, No. 5, December 1990, pp. 435–450.

[46] P. Comon, C. Jutten and J. Hérault, "Blind Separation of Sources, Part II: Problems Statement", *Signal Processing*, Vol. 24, No. 1, July 1991, pp. 11–20.

[47] P. Comon, "MA Identification Using Fourth Order Cumulants", *Signal Processing*, Vol. 26, 1992, pp. 381–388.

[48] P. Comon, "Independent Component Analysis, A New Concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 287–314.

[49] P. Comon, "Tensor Diagonalization, A Useful Tool in Signal Processing", in: *Proceedings IFAC SYSID'94*, Copenhagen, July 1994, pp. 77–82.

[50] P. Comon, "Contrasts for Multichannel Blind Deconvolution", *IEEE Signal Processing Letters*, Vol. 3, No. 7, July 1996, pp. 209–211.

[51] P. Comon and B. Mourrain, "Decomposition of Quantics in Sums of Powers of Linear Forms", *Signal Processing*, Vol. 53, No. 2, September 1996, pp. 93–107.

[52] P. Comon and E. Moreau, "Improved Contrast Dedicated to Blind Separation in Communications", in: *Proceedings ICASSP-97, IEEE International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, April 20–24, 1997, pp. 3453–3456.

[53] P. Comon and P. Chevalier, "Blind Source Separation: Models, Concepts, Algorithms and Performance", in: S. S. Haykin (Ed.) *Unsupervised Adaptive Filtering*, Wiley Series in Adaptive and Learning Systems for Communications, Signal Processing and Control, New York: John Wiley & Sons, to appear in December 1999.

[54] P. Comon and O. Grellier, "Non Linear Inversion of Underdetermined Mixtures", *Proceedings ICA'99, First International Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, Jan. 11–15, 1999, pp. 461–465.

[55] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, 1991.

[56] N. Delfosse and P. Loubaton, "Adaptive Blind Separation of Independent Sources: A Deflation Approach", *Signal Processing*, Vol. 45, 1995, pp. 59–83.

[57] L. Downs, "Quaternions References for CS184 [Foundations of Computer Graphics]", University of California, Berkeley, USA, URL:

`http://www.cs.berkeley.edu/~laura/cs184/quat` [updated October 9, 1997; visited November 1998].

[58] M. W. Eysenck (Ed.), *The Blackwell Dictionary of Cognitive Psychology*, Oxford: Blackwell, 1990.

[59] M. W. Eysenck and M. T. Keane, *Cognitive Psychology: A Student's Handbook*, Hove, UK: Lawrence Erlbaum Associates Ltd., 1990.

[60] M. W. Eysenck, *Principles of Cognitive Psychology*, Hove, UK: Lawrence Erlbaum Associates Ltd., 1993.

[61] P. Fabry, Ch. Servière and J.-L. Lacoume, "Improving Signal Subspace Estimation and Source Number Detection in the Context of Spatially Correlated Noises", in: *Proceedings EUSIPCO-98, IX European Signal Processing Conference*, Rhodes, Greece, September 8–11, 1998, Vol. III, pp. 1809–1812.

[62] M. Gaeta and J.-L. Lacoume, "Sources Separation without a priori Knowledge: the Maximum Likelihood Solution", in: L. Torres *et al.* (Eds.), *Proceedings EUSIPCO-90, V European Signal Processing Conference*, Barcelona, Spain, 1990, pp. 621–624.

[63] J. A. Gallian, *Contemporary Abstract Algebra*, Fourth Edition, Boston, Massachussets: Houghton Mifflin Company, 1998.

[64] G.-O. Glentis, K. Berberidis and S. Theodoridis, "Efficient Least Squares Adaptive Algorithms for FIR Transversal Filtering", *IEEE Signal Processing Magazine*, Vol. 16, No. 4, July 1999, pp. 13–41.

[65] V. P. Godambe (Ed.), *Estimating Functions*, Oxford Statistical Science Series, Oxford: Oxford University Press, 1991.

[66] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Second Edition, Baltimore, Maryland: The Johns Hopkins University Press, 1989.

[67] W. R. Hamilton, *Elements of Quaternions*, London: Longmans, Green & Co., 1866.

[68] F. Harroy, J.-L. Lacoume and M. A. Lagunas, "A General Adaptive Algorithm for nonGaussian Source Separation without Any Constraint", in: M. Holt *et al.* (Eds.), *Proceedings EUSIPCO-94, VII European Signal Processing Conference*, Edinburgh, Scotland, UK, September 1994, pp. 1161–1164.

[69] F. Harroy and J.-L. Lacoume, "Maximum Likelihood Estimators and Cramer-Rao Bounds in Source Separation", *Signal Processing*, Vol. 55, No. 2, December 1996, pp. 167–177.

[70] S. S. Haykin, *Adaptive Filter Theory*, Third Edition, Upper Saddle River, New Jersey: Prentice Hall, 1996.

[71] J. Hérault, C. Jutten and B. Ans, "Détection de Grandeurs Primitives dans un Message Composite par une Architecture Neuromimétique en Apprentissage Non Supervisé", *Proceedings 10th GRETSI Colloquium*, Nice, France, May 20–24, 1985, pp. 1017–1022.

[72] F. Herrmann, *Personal Communication*, Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, UK, August 1999.

[73] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Second Edition, Prentice Hall, 1988.

[74] C. Jutten and J. Hérault, "Une Solution Neuromimétique au Problème de Séparation de Sources", *Traitement du Signal*, Vol. 5, No. 6, 1989, pp. 389–404.

[75] C. Jutten and J. Hérault, "Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture", *Signal Processing*, Vol. 24, No. 1, July 1991, pp. 1–10.

[76] J. Karhunen, "Neural Approaches to Independent Component Analysis and Source Separation", *Proceedings ESANN'96, 4th European Symposium on Artificial Neural Networks*, Bruges, Belgium, April 24–26, 1996.

[77] P. Kelland and P. G. Tait, *Introduction to Quaternions*, Second Edition, London: MacMillan & Co., 1882.

[78] S. Kotz and N. L. Johnson, *Encyclopedia of Statistical Sciences*, New York: John Wiley & Sons, 1982.

[79] J.-L. Lacoume and P. Ruiz, "Sources Identification: A Solution Based on the Cumulants", *Proceedings ICASSP-88, IEEE International Conference on Acoustics, Speech and Signal Processing*, Minneapolis, USA, August 3–5, 1988.

[80] J.-L. Lacoume, P.-O. Amblard et P. Comon, *Statistiques d'Ordre Supérieur pour le Traitement du Signal*, Collection Sciences de l'Ingénieur, Paris: Masson, 1997.

[81] J.-L. Lacoume, M. Gaeta and P.-O. Amblard, "From Order 2 to HOS: New Tools and Applications", in: J. Vandewalle *et al.* (Eds.), *Proceedings EUSIPCO-92, VI European Signal Processing Conference*, Brussels, 1992, pp. 91–98.

[82] J.-L. Lacoume and P. Ruiz, "Separation of Independent Sources from Correlated Inputs", *IEEE Transactions on Signal Processing*, Vol. 40, No. 12, December 1992, pp. 3074–3078.

[83] J.-L. Lacoume, F. Glangeaud and J. Mars, "Blind Separation of Polarised Waves", in: *Proceedings EUSIPCO-98, IX European Signal Processing Conference*, Rhodes, Greece, September 8–11, 1998, Vol. III, pp. 1629–1632.

[84] B. H. Laheld and J.-F. Cardoso, "Adaptive source separation with uniform performance", in: M. Holt *et al.* (Eds.), *Proceedings EUSIPCO-94, VII European Signal Processing Conference*, Edinburgh, Scotland, UK, September 1994, pp. 183–186.

[85] L. De Lathauwer, B. De Moor and J. Vandewalle, "Blind Source Separation by Higher-Order Singular Value Decomposition", in: M. Holt *et al.* (Eds.), *Proceedings EUSIPCO-94, VII European Signal Processing Conference*, Edinburgh, Scotland, UK, September 1994, pp. 175–178.

[86] L. De Lathauwer, D. Callaerts, B. De Moor and J. Vandewalle, "Fetal Electrocardiogram Extraction by Source Subspace Separation", in: *Proceedings IEEE/ATHOS Signal Processing Conference on Higher-Order Statistics*, Girona, Spain, June 12–14, 1995, pp. 134–138.

[87] L. De Lathauwer, B. De Moor and J. Vandewalle, "Blind Source Separation by Simultaneous Third-Order Tensor Diagonalization", in: *Proceedings EUSIPCO-96, VIII European Signal Processing Conference*, Trieste, Italy, September 10-13, 1996, pp. 2089–2092.

[88] A. Mansour and C. Jutten, "Fourth-Order Criteria for Blind Sources Separation", *IEEE Transactions on Signal Processing*, Vol. 43, No. 8, August 1995, pp. 2022–2025.

[89] A. Mansour and C. Jutten, "A Direct Solution for Blind Separation of Sources", *IEEE Transactions on Signal Processing*, Vol. 44, No. 3, March 1996, pp. 746–748.

[90] K. V. Mardia, J. T. Tent and J. M. Bibby, *Multivariate Analysis*, London: Academic Press, 1979.

[91] P. McCullagh, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1987.

[92] J. M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *Proceedings of the IEEE*, Vol. 79, No. 3, March 1991, pp. 278–305.

[93] B. De Moor (Ed.), "DaISy: Database for the Identification of Systems", ESAT/SISTA, K. U. Leuven, Belgium, `URL:` `http://www.esat.kuleuven.ac.be/sista/daisy`, dataset: Biomedical Systems [96-012], May 1999.

[94] E. Moreau, "Criteria for Complex Source Separation", in: *Proceedings EUSIPCO-96, VIII European Signal Processing Conference*, Trieste, Italy, September 10–13, 1996, pp. 931–934.

[95] E. Moreau and O. Macchi, "High-Order Contrasts for Self-Adaptive Source Separation", *International Journal of Adaptive Control and Signal Processing*, Vol. 10, 1996, pp. 19–46.

[96] A. K. Nandi and V. Zarzoso, "Fourth-Order Cumulant Based Blind Source

Separation", *IEEE Signal Processing Letters*, Vol. 3, No. 12, December 1996, pp. 312–314.

[97] A. K. Nandi and V. Zarzoso, "Foetal ECG Separation", *IEE Colloquium on the Use of Model Based Digital Signal Processing Techniques in the Analysis of Biomedical Signals*, Savoy Place, London, April 16, 1997, No. 9, pp. 8/1–8/6.

[98] A. K. Nandi (Ed.), *Blind Estimation Using Higher-Order Statistics*, Boston: Kluwer Academic Publishers, 1999.

[99] C. L. Nikias and M. R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework", *Proceedings of the IEEE*, Vol. 75, No. 7, July 1987, pp. 869–891.

[100] C. L. Nikias and J. M. Mendel, "Signal Processing with Higher-Order Statistics", *IEEE Signal Processing Magazine*, Vol. 10, No. 3, July 1993, pp. 10–37.

[101] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*, Englewood Cliffs, New Jersey: Prentice Hall, Signal Processing Series, 1993.

[102] T. Oostendorp, *Modelling the Fetal ECG*, PhD Thesis, K. U. Nijmegen, The Netherlands, 1989.

[103] D. T. Pham, P. Garat and C. Jutten, "Separation of a Mixture of Independent Sources through a Maximum-Likelihood Approach", in: J. Vandewalle *et al.* (Eds.), *Proceedings EUSIPCO-92, VI European Signal Processing Conference*, Brussels, 1992, pp. 771–774.

[104] D. T. Pham, "Blind Separation of Instantaneous Mixture of Sources via an Independent Component Analysis", *IEEE Transactions on Signal Processing*, Vol. 44, No. 11, November 1996, pp. 2768–2779.

[105] R. Plonsey, *Bioelectric Phenomena*, New York: McGraw-Hill, 1969.

[106] R. Plonsey and R. C. Barr, *Bioelectricity: A Quantitative Approach*, New York: Plenum Press, 1988.

[107] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge: Cambridge University Press, 1988.

[108] J. G. Proakis, *Digital Communications*, Second Edition, McGraw-Hill Series in Electrical Engineering, New York: McGraw-Hill, 1989.

[109] R. Roy and T. Kailath, "ESPRIT — Estimation of Signal Parameters via Rotational Invariance Techniques", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 7, July 1989, pp. 984–995.

[110] H. Sahlin, *Blind Signal Separation by Second Order Statistics*, PhD Thesis, Chalmers University of Technology, Göteborg, Sweden, October 1998.

[111] L. L. Scharf, *Statistical Signal Processing. Detection, Estimation and Time Series Analysis*, Addison-Wesley Series in Electrical and Computer Engineering: Digital Signal Processing, Reading, Massachusetts: Addison-Wesley, 1991.

[112] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Transactions on Antennas and Propagation*, Vol. AP-34, No. 3, March 1986, pp. 276–280.

[113] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons, 1980.

[114] O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)", *IEEE Transactions on Signal Processing*, Vol. 36, No. 2, March 1990, pp. 312–321.

[115] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986.

[116] V. C. Soon, L. Tong, Y. F. Huang and R. Liu, "An Extended Fourth Order Blind Identification Algorithm in Spatially Correlated Noise", *Proceedings ICASSP-90, IEEE International Conference on Acoustics, Speech and Signal Processing*, Albuquerque, New Mexico, USA, April 3–6, 1990, Vol. 3, pp. 1365–1368.

[117] E. Sorouchyari, "Blind separation of sources, Part III: Stability analysis", *Signal Processing*, Vol. 24, No. 1, July 1991, pp. 21–29.

[118] G. Spence, "Lower Order Radial Weighting Techniques for Improving the Robustness of Blind Signal Separation Methods", in: *IEE Colloquium on Statistical Signal Processing*, Savoy Place, London, January 6, 1999, No. 2, pp. 4/1–4/6.

[119] P. Stoica, "Asymptomania?", *IEEE Signal Processing Magazine*, Vol. 15, No. 1, January 1998, p. 16.

[120] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. I, Sixth Edition, London: Edward Arnold, 1994.

[121] A. Swami, G. B. Giannakis and G. Zhou, "Bibliography on Higher-Order Statistics", *Signal Processing*, Vol. 60, No. 1, July 1997, pp. 65–126.

[122] L. Tong, R. Liu, V. C. Soon and Y.-F. Huang, "Indeterminacy and Identifiability of Blind Identification", *IEEE Transactions on Signal Processing*, Vol. 38, No. 5, May 1991, pp. 499–509.

[123] L. Tong, S. Yu, Y. Inouye and R. Liu, "A Necessary and Sufficient Condition of Blind Identification", in: *Proceedings of the International Signal Processing Workshop on Higher-Order Statistics*, Chamrousse, France, July 10–12, 1991, pp. 261–264.

[124] L. Tong, Y. Inouye and R. Liu, "Waveform-Preserving Blind Estimation of Multiple Independent Sources", *IEEE Transactions on Signal Processing*, Vol. 41, No. 7, July 1993, pp. 2461–2470.

[125] J. Vanderschoot, D. Callaerts, W. Sansen, J. Vandewalle, G. Vantrappen and J. Janssens, "Two Methods for Optimal MECG Elimination and FECG Detection from Skin Electrode Signals", *IEEE Transactions on Biomedical Engineering*, Vol. BME-34, No. 3, March 1987, pp. 233–243.

[126] D. L. Wallace, "Asymptotic Approximations to Distributions", *Annals of Mathematical Statistics*, Vol. 29, 1958, pp. 635–654.

[127] R. J. Webster, "An Efficient Source of Random Numbers for Modeling Symmetrically Distributed Noise", *IEEE Transactions on Signal Processing*, Vol. 44, No. 1, January 1996, pp. 147–150.

[128] E. Weinstein, M. Feder and A. V. Oppenheim, "Multi-Channel Signal Separation by Decorrelation", *IEEE Transactions on Speech and Audio Processing*, Vol. 1, No. 4, October 1993, pp. 405–413.

[129] B. Widrow, J. R. Glover, J. M. McCool *et al.*, "Adaptive Noise Cancelling: Principles and Applications", *Proceedings of the IEEE*, Vol. 63, No. 12, December 1975, pp. 1692–1716.

[130] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, New Jersey: Prentice Hall, 1985.

[131] D. Yellin and E. Weinstein, "Criteria for Multichannel Signal Separation", *IEEE Transactions on Signal Processing*, Vol. 42, No. 8, August 1994, pp. 2158–2168.

[132] D. Yellin and E. Weinstein, "Multichannel Signal Separation: Methods and Analysis", *IEEE Transactions on Signal Processing*, Vol. 44, No. 1, January 1996, pp. 106–118.

[133] V. Zarzoso and A. K. Nandi, "The Potential of Decorrelation in Blind Separation of Sources Based on Cumulants", in: *Proceedings ECSAP-97, First European Conference on Signal Analysis and Prediction*, Prague, Czech Republic, June 24–27, 1997, pp. 293–296.

[134] V. Zarzoso, A. K. Nandi and E. Bacharakis, "Maternal and Foetal ECG Separation Using Blind Source Separation Methods", *IMA Journal of Mathematics Applied in Medicine & Biology*, Vol. 14, No. 3, 1997, pp. 207–225.

[135] V. Zarzoso and A. K. Nandi. "Modelling Signals of Arbitrary Kurtosis for Testing BSS Methods", *IEE Electronics Letters*, Vol. 34, No. 1, January 8, 1998, pp. 29–30. (Errata: Vol. 34, No. 7, April 2, 1998, p. 703.)

[136] V. Zarzoso y A. K. Nandi, "Extracción del ECG fetal mediante técnicas de separación ciega de fuentes", en: *Actas CASEIB'98, XVI Congreso Anual de la*

Sociedad Española de Ingenieros Biomédicos, Valencia, 21–23 de Septiembre de 1998, pp. 89–92 [Spanish].

[137] V. Zarzoso and A. K. Nandi, "Blind Source Separation without Optimization Criteria?", in: *Proceedings ICASSP-99, IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, Arizona, USA, March 15–19, 1999, Vol. III, pp. 1453–1456.

[138] V. Zarzoso and A. K. Nandi, "Blind Source Separation", in: A. K. Nandi (Ed.), *Blind Estimation Using Higher-Order Statistics*, Boston: Kluwer Academic Publishers, 1999, pp. 167–252.

[139] V. Zarzoso and A. K. Nandi, "Blind Separation of Independent Sources for Virtually Any Source Probability Density Function", *IEEE Transactions on Signal Processing*, Vol. 47, No. 9, September 1999, pp. 2419–2432.

[140] V. Zarzoso and A. K. Nandi, "Non-Invasive Fetal Electrocardiogram Extraction: Blind Separation vs. Adaptive Noise Cancellation", submitted to *IEEE Transactions on Biomedical Engineering*, May 1999.

[141] V. Zarzoso and A. K. Nandi, "Analysis and Development of Closed-Form Estimators for Blind Separation of Sources — Part II: Complex Mixtures", submitted to *IEEE Transactions on Signal Processing*, June 1999.