

Helsinki University of Technology Signal Processing Laboratory

Teknillinen korkeakoulu Signaalinkäsittelytekniikan laboratorio

Espoo 2001

Report 31

ARRAY AND MULTICHANNEL SIGNAL PROCESSING USING NONPARAMETRIC STATISTICS

Samuli Visuri

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4 at Helsinki University of Technology (Espoo, Finland) on the 2nd of March, 2001, at 12 o'clock noon.

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

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Signaalinkäsittelytekniikan laboratorio

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

© Samuli Visuri

ISBN 951-22-5344-5
ISSN 1456-6907

Otamedia Oy
Espoo 2001

Abstract

In array signal processing a group of sensors located at distinct spatial locations is deployed to measure a propagating wavefield. The multichannel output is then processed to provide information about parameters of interest. Application areas include smart antennas in communications, radar, sonar and biomedicine.

When deriving array signal processing algorithms the noise is typically modeled as a white Gaussian random process. A shortcoming of the estimation procedures derived under Gaussian assumption is that they are extremely sensitive to deviations from the assumed model, i.e. they are not robust. In real-world applications the assumption of white Gaussian noise is not always valid. Consequently, there has been a growing interest in estimation methods which work reliably in both Gaussian and non-Gaussian noise.

In this thesis, new statistical procedures for array and multichannel signal processing are developed. In the area of array signal processing, the work concentrates on high-resolution subspace-based Direction Of Arrival (DOA) estimation and estimation of the number of source signals. Robust methods for DOA estimation and estimation of the number of source signals are derived. Spatial-smoothing based extensions of the techniques to deal with coherent signals are also derived. The methods developed are based on multivariate nonparametric statistics, in particular sign and rank covariance matrices. It is shown that these statistics may be used to obtain convergent estimates of the signal and noise subspaces for a large family of symmetric noise distributions. Simulations reveal that the techniques developed exhibit near-optimal performance when the noise distribution is Gaussian and are highly reliable if the noise is non-Gaussian.

Multivariate nonparametric statistics are also applied to frequency estimation and estimation of the eigenvectors of the covariance matrix. Theoretical justification for the techniques is shown and their robust performance is illustrated in simulations.

Preface

The research work for this thesis was carried out at the Signal Processing Laboratory, Helsinki University of Technology and the Signal Processing Laboratory, Tampere University of Technology, during the years 1998-2000. I wish to express my gratitude to my co-supervisors, Prof. Visa Koivunen, Helsinki University of Technology and Prof. Hannu Oja, University of Jyväskylä for their expert guidance and constant support. A special thanks to Visa for making it a pleasure to work in his research group.

I would like to thank my thesis reviewers Prof. Jukka Nyblom and Prof. Saleem Kassam for their constructive comments which truly helped in the clarification of the presentation. To Dr. Charles Murphy I am in great debt for revising the language. I am also grateful to my colleagues and co-workers, especially Mihai Enescu, Jan Eriksson, Dr. Jukka Mannerkoski, Dr. Jyrki Möttönen and Esa Ollila for our many interesting discussions, some of which even concerned science.

This research was funded by the Academy of Finland, the Finnish Defence Forces Research Institute of Technology and the Graduate School in Electronics Telecommunications and Automation (GETA). Additional financial support for my doctoral studies was provided by the Emil Aaltonen Foundation, the Nokia Foundation and the Seppo Säynäjäkangas Science Foundation. These organizations and foundations are greatly acknowledged for making this work possible.

In addition, I would like to thank my parents Marja-Liisa and Olavi for teaching me what is valuable in life and for supporting me in all my goals. My brother Tuomas as well as my grandparents also deserve a warm thanks. Furthermore, my thanks go to all my friends, especially Manne and Simo, for their positive contribution to my leisure time

activities.

Finally, I would like to thank my lovely fiancée Minna for her patience, love and support.

Otaniemi, Espoo

January 30, 2001

Samuli Visuri

Contents

Abstract	i
Preface	iii
List of original publications	vii
List of abbreviations and symbols	ix
1 Introduction	1
1.1 Motivation	1
1.2 Scope of the thesis	3
1.3 Contributions of the thesis	4
1.4 Summary of publications	5
2 Array signal processing	7
2.1 Signal model	8
2.2 Review of DOA estimation methods	10
2.2.1 The Conventional beamformer	11
2.2.2 Capon's method	12
2.2.3 Subspace methods	14
2.2.4 Maximum likelihood methods	22
2.2.5 Subspace fitting methods	25
2.3 Estimation of the number of signals	27
2.4 Discussion	30

3	Robust estimation of the covariance matrix	31
3.1	Tools for measuring robustness	32
3.1.1	Breakdown point	33
3.1.2	Influence function	35
3.2	A review of robust estimators of the covariance matrix	37
3.2.1	M -estimation	39
3.2.2	MVE and MCD	41
3.2.3	S -estimates	42
3.2.4	Estimates based on projections	43
3.3	Robust estimation using nonparametric statistics	45
3.4	Discussion	50
4	Robust DOA estimation	53
4.1	Noise distribution families	53
4.2	Review of robust DOA estimation methods	55
4.3	Robust estimation using nonparametric statistics	56
4.3.1	DOA estimation	57
4.3.2	Estimating the number of sources	59
4.3.3	Coherent signals	60
4.4	Simulation results	62
4.4.1	Non-coherent signals in noise	62
4.4.2	Coherent signals in noise	64
4.5	Frequency estimation using nonparametric statistics	66
4.6	Discussion	70
5	Summary	71
A	Proof of Theorem 1	73
	Bibliography	75

List of original publications

I

S. Visuri, V. Koivunen and H. Oja. Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, vol. 91, no. 2, pp. 557–575, 2000.

II

S. Visuri, H. Oja and V. Koivunen. Multichannel signal processing using spatial rank covariance matrices. In *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99)*, vol. 1, pp. 75–79, (Antalya, Turkey), June 1999.

III

S. Visuri, H. Oja and V. Koivunen. Robust subspace DoA estimation for wireless communications. In *Proceedings of the 2000 IEEE 51st Vehicular Technology Conference (VTC2000-Spring)*, vol. 3, pp. 2551–2555, (Tokyo, Japan), May 2000.

IV

S. Visuri, H. Oja and V. Koivunen. Nonparametric statistics for DOA estimation in the presence of multipath. In *Proceedings of the First IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2000)*, pp. 356–360, (Cambridge, MA, USA), March 2000.

V

S. Visuri, H. Oja and V. Koivunen. Subspace-based direction of arrival estimation using nonparametric statistics. Helsinki University of Technology Signal Processing Laboratory, Report 30, 2000. *IEEE Transactions on Signal Processing*, accepted for publication with minor revision.

VI

S. Visuri, H. Oja and V. Koivunen. Nonparametric statistics for subspace based frequency estimation. In *Proceedings of the X European Signal Processing Conference (EUSIPCO-2000)*, vol. 3, pp. 1261–1264, (Tampere, Finland), September 2000.

List of abbreviations and symbols

Abbreviations

AIC	Akaike information criterion	RCM	rank covariance matrix
AMS	autocorrelation matrix smoothing	RGB	red green blue
AR	autoregressive	QAM	quadrature amplitude modulation
BSS	blind source separation		
DML	deterministic maximum likelihood	SAM	sign autocovariance matrix
DOA	direction of arrival	SCM	sign covariance matrix
EM	expectation maximization	SML	stochastic maximum likelihood
LS	least-squares	SNR	signal-to-noise ratio
MAD	median absolute deviation	SSF	signal subspace fitting
ML	maximum likelihood	TCM	tau covariance matrix
MCD	minimum covariance determinant	TLS	total least squares
MVE	minimum volume ellipsoid	ULA	uniform linear array
NSF	noise subspace fitting	WSF	weighted subspace fitting

Symbols

a^*	complex conjugate of the scalar a
\mathbf{a}^H	Hermitian transpose of the vector \mathbf{a}
A^T	transpose of the matrix A
A^\dagger	pseudo-inverse of A
$A(\boldsymbol{\theta})$	matrix of steering vector related to the DOAs $\boldsymbol{\theta}$
$\mathbf{a}(\theta)$	steering vector related to the DOA θ
d	element spacing between the sensors
\det	determinant
j	imaginary unit
K	number of signals, number of complex exponentials
K'	rank of the signal covariance matrix
L	number of subarrays
med	median
M	number of sensors, dimension of the data
N	number of observations
P	size of a subarray
$PDS(M)$	set of all $M \times M$ positive definite symmetric matrices
$\mathbf{R}(\cdot)$	spatial rank function
R_S	sample sign autocovariance matrix
$s_k(n)$	k th signal component
$\mathbf{s}(n)$	signal vector
$\mathbf{S}(\cdot)$	spatial sign function
S	sample covariance matrix
S_1	sample sign covariance matrix
S_2	sample tau covariance matrix
S_3	sample rank covariance matrix
Tr	trace

\mathbf{u}_m	eigenvector corresponding to the m th largest eigenvalue
U	matrix of the eigenvectors
U_s	matrix of the signal subspace eigenvectors
U_n	matrix of the noise subspace eigenvectors
$\mathbf{v}(n)$	additive noise
$\mathbf{v}_l^f(n)$	additive noise in the l th forward subarray
$\mathbf{v}_l^b(n)$	additive noise in the l th backward subarray
$\mathbf{x}(n)$	array output vector
\mathbf{x}_l^b	received signal at the l th backward subarray
\mathbf{x}_l^f	received signal at the l th forward subarray
δ_{lk}	Kronecker delta
λ_m	m th largest eigenvalue
Π_A	projection matrix to the signal subspace
Π_A^\perp	projection matrix to the noise subspace
σ^2	noise variance
σ_s^2	signal power
Σ	covariance matrix
Σ_1	sign covariance matrix
Σ_2	tau covariance matrix
Σ_3	rank covariance matrix
Σ_s	signal covariance matrix
θ_k	DOA of the k th signal
$\boldsymbol{\theta}$	vector of DOAs
$ \cdot $	modulus
$\ \cdot\ $	Euclidean norm
$\lfloor a \rfloor$	nearest integer less than or equal to a
$\#$	number of elements

Chapter 1

Introduction

1.1 Motivation

In array signal processing a group of sensors located at distinct spatial locations is deployed to measure a propagating wavefield. The multichannel output is then processed to provide information about parameters of interest. The use of antenna arrays is one of the key features of future wireless communication systems. For example, an adaptive array (smart antenna) may be used to provide high gain in the direction of a desired transmitter while steering nulls in the direction of interferers. The benefit obtained is an increase in signal-to-interference-and-noise-ratio resulting to higher system capacity. In radar applications an array of active sensors radiates electromagnetic pulses and measures return signals. The radar returns enable estimation of velocity (Doppler frequency), range and Direction Of Arrival (DOA). In sonar, arrays of hydrophones measure acoustic signals in order to detect and locate distant sources. In biomedicine array signal processing methods are used, for example, to localize brain activity using biomagnetic sensor arrays.

The physical measurements collected by a sensor array contain noise. When deriving array signal processing algorithms the noise is conventionally modeled as a white Gaussian random process. The Gaussian (normal) distribution is the most often used probability distribution in statistical signal processing. The distribution is related to the Least-Squares (LS) estimation method introduced by Legendre and Gauss in the 19th century. It is the

error distribution for which the LS method is optimal.

One reason for the popularity of the Gaussian assumption is of course that its use can be motivated by the central limit theorem. Another reason is undoubtedly the fact that under a Gaussian assumption, the derivation of optimal estimators and analysis of their properties are straightforward, as was noted by Gauss himself:

This idea, however, from its nature, involves something vague ... and clearly innumerable different principles can be proposed ... But of all these principles ours is the most simple; by the others we shall be led into the most complicated calculations.

Gauss in 1809, on the least squares criterion¹

The weakness of the optimal estimation procedures derived under Gaussian assumption is that they are extremely sensitive to deviations from the assumed model, i.e. they are not robust. In many signal processing applications the assumption of white Gaussian noise is not always completely satisfied. For example, it has been observed through experimental measurements that the ambient noise in indoor and urban radio channels is decidedly non-Gaussian (c.f. [69]). Consequently, there has been a growing interest in estimation algorithms which work properly both in Gaussian and non-Gaussian noise environments [47, 81, 115, 54].

In most of the array signal processing applications the first task is to estimate the DOAs of the incoming signals. All DOA estimation algorithms need information about the number of source signals. If this information is not provided, it has to be estimated from the data. Most DOA estimation algorithms and methods for estimating the number of signals are based on the array output covariance matrix. In the algorithms the covariance matrix or its eigenvalues or eigenvectors are used to provide the information needed for the estimation task. The standard estimator for the array output covariance matrix is the sample covariance matrix. Use of the sample covariance matrix may cause the estimation methods to produce unreliable estimates if the noise is non-Gaussian. This is because the sample covariance matrix is an optimal estimator for Gaussian data and hence not a robust

¹Quotation taken from [92].

estimator. If the necessary information (covariance matrix or its eigenvectors or eigenvalues) could be estimated robustly, it could be used in various DOA estimation algorithms as well as in estimation of the number of signals. In other words, the conventional methods may be made more robust by replacing the sample covariance matrix with an robust estimator of the covariance matrix, or by estimating the eigenvectors or eigenvalues of the array output covariance matrix in a robust manner.

Also a large number of signal processing applications which are not explicitly based on arrays require processing of multichannel data and consequently estimation of covariance matrices. The application domains include biomedical signal processing such as EEG, image analysis and color image processing. Robust covariance matrix estimators may also be used in these applications when the Gaussian assumption of the data does not hold. In addition, robust covariance matrix estimators may be applied to conventional multivariate statistical analysis.

1.2 Scope of the thesis

The scope of this thesis is to develop new procedures for array and multichannel signal processing and develop tools for characterizing robustness of matrix valued statistics. In array signal processing the work is limited to subspace-based DOA estimation and estimation of the number of source signals.

The design goal of the estimation techniques is robustness against heavy-tailed non-Gaussian noise of the type appearing in many real world applications. In case of DOA estimation the methods should have high resolution, i.e. they should be able to distinguish between one source and two sources with close DOAs. The computational complexity of the resulting algorithms should not be too high when compared to existing algorithms developed for Gaussian noise. Also the performance of the methods should be near-optimal when the noise distribution is Gaussian. Rigorous mathematical theory should buttress the derived methods.

The characterization of robustness of a covariance matrix estimator is conventionally

done by using only its eigenvalues. Since the eigenvectors of a covariance matrix are crucial in many applications there is a need to develop tools that take into account also the eigenvectors. These tools may then be used in describing both quantitative and qualitative robustness.

1.3 Contributions of the thesis

The new estimation techniques developed in this thesis are based on multivariate nonparametric statistics, in particular spatial sign and rank covariance matrices. It is shown that the Sign Covariance Matrix (SCM) and the Tau Covariance Matrix (TCM) contain enough information to estimate the eigenvectors of the covariance matrix for a large family of symmetric distributions. Moreover, it is shown that the estimates of the SCM and the TCM are convergent, i.e. they converge to the correct values when the number of data samples N tends to infinity. New tools for analyzing robustness of covariance matrix estimators are developed. The tools are based on the eigendecomposition of a covariance matrix [5].

Robust algorithms for DOA estimation and estimation of the number of source signals are derived. Spatial-smoothing based extensions of the algorithms to deal with coherent signals are derived as well. Theoretical motivation of the algorithms is shown for a large family of noise distributions. It is shown using simulations that the resulting techniques perform reliably regardless of the heavy-tailed nature of the noise distribution.

Multivariate nonparametric statistics are also applied to estimate the frequencies of complex exponentials from time-series data. It is shown that for Gaussian noise the methods produce convergent estimates. The efficiency and robust performance of the techniques in non-Gaussian noise is shown using simulations. The use of spatial sign and rank covariance matrices is demonstrated in several multichannel signal processing problems including RGB color image filtering.

1.4 Summary of publications

The remainder of this thesis is organized as follows. Chapter 2 introduces the common signal model and the basic concepts employed in array signal processing. A brief review of widely used direction of arrival estimation techniques is given and the problem of estimating the number of signals is addressed. Chapter 3 contains a review of robust covariance matrix estimation techniques and introduces covariance matrix estimation techniques based on multivariate nonparametric statistics. In addition, perturbation analysis of covariance matrix estimators is discussed. In chapter 4, robust DOA estimation techniques are reviewed and new robust methods are introduced. The estimation of the number of source signals is considered as well. The closely related problem of frequency estimation is introduced and robust methods developed. The techniques considered are robust against heavy-tailed or impulsive non-Gaussian noise. Finally, chapter 5 concludes the thesis.

This thesis consists of 6 publications on robust DOA estimation, robust estimation of the number of sources, and multichannel signal processing. In paper *I*, tools analyzing the distinct properties of covariance matrix estimators are introduced. These tools include sensitivity plots that characterize qualitative robustness in a fashion similar to that of the influence function. Three different concepts of sign and rank covariance matrices are discussed and their use in covariance matrix estimation is proposed. Qualitative and quantitative robustness as well as finite sample efficiencies of the two proposed methods are studied. A signal processing example where robust covariance matrix estimates are needed is considered as well. Paper *II* demonstrates the use of spatial rank covariance matrices in different multichannel signal processing tasks. Robust covariance matrix estimates obtained from the sample Rank Covariance Matrix (RCM) and the sample TCM are used in RGB color image filtering, principal component analysis, discrete Karhunen-Loève transform and Blind Source Separation (BSS). In addition, it is shown how methods based on the sample covariance matrix give strongly misleading results in the face of outliers.

In paper *III*, the definitions of the RCM and the TCM are extended for complex-valued data. The use of the sample RCM and the sample TCM is proposed in the DOA estimation. The algorithms are based on estimating the signal or noise subspace from these

nonparametric statistics. Reliable performance of the methods is shown using simulations of Gaussian and non-Gaussian noise conditions. DOA estimation of coherent signals is addressed in paper *IV*. The techniques are based on the SCM and the TCM. Spatial smoothing is employed as a preprocessing step in order to deal with coherent sources. Theoretical motivation for the resulting algorithms is shown. The performance of the algorithms is studied using simulations. The results show that near-optimal performance is obtained in a wide variety of different noise conditions.

Paper *V* is the main publication of this thesis. A proof is presented that the sample SCM and the sample TCM can be used to obtain convergent estimates of the signal and noise subspaces. These estimates are then used in DOA estimation. It also proves that the sample SCM and TCM converge with probability one to the corresponding theoretical matrices. In addition, the important problem of estimating the number of signals is considered and various simulation results are reported.

Paper *VI* introduces sample Sign Autocovariance Matrix (SAM) and discusses its use in frequency estimation. It is proven that when the noise is circular Gaussian, the sample SAM can be used to estimate the signal and noise subspaces spanned by the eigenvectors of the autocovariance matrix. A robust method for estimating the number of complex exponentials is proposed.

All of the simulation software for all of the original papers of this dissertation was written by the author, with the exception of that of the BSS algorithm used in *I* and *II*. The author participated in writing and planning experiments for paper *I*. The author derived the analytical results and the algorithms in and did most of the writing of papers *II-VI*. The coauthors collaborated in experiment design, provided guidance for the author's proofs, and contributed to the final version of each paper.

Chapter 2

Array signal processing

In array signal processing a group of sensors located at distinct spatial locations is deployed to measure a propagating wavefield. The multivariate output is then processed to provide information about parameters of interest. Application areas which use arrays include communications, radar, sonar, seismology, biomedicine and astronomy.

In most of the applications the first task is to estimate the Directions Of Arrival (DOAs) of the incoming signals. This information can then be used to localize the signal sources, form high gains to the desired DOAs or to steer nulls into direction of interferers. All DOA estimation algorithms need information about the number of source signals. If this information is not provided, it has to be estimated from the data. DOA estimation and estimation of the number of signals are key issues of array signal processing. This chapter focuses on these concepts and discusses differences among the reviewed estimation methods.

There are several ways to compare different estimators. The methods should be *consistent*, i.e. the estimates should converge to the correct values when the number of observations tends to infinity. The efficiency of an estimator is measured by its variance. For unbiased estimators the Cramér-Rao bound can be used to compare efficiency of different estimators. Resolution, ability to distinguish between one source and two sources with close DOAs, is also an important property. A desirable estimator should be able to distinguish between sources with arbitrarily close DOAs. Also computational complexity is a concern. A goal of estimator design is to have high resolution and efficiency at low computational

cost.

The chapter introduces the common signal model and the basic concepts employed in array signal processing. A brief review of widely used DOA estimation techniques is given and estimation of the number of signals is addressed. The chapter ends with discussion.

2.1 Signal model

The development of the signal model is based on the number of simplifying assumptions. The sources are assumed to be *narrow band* and situated in the *far field* of the array. Furthermore, we assume that both the sources and the sensors in the array are in the *same plane* and that the sources are *point emitters*. In addition, it is assumed that the *propagation medium is homogeneous (i.e. not dispersive)*. Consequently, the waves arriving at the array can be considered to be *planar*. Under these assumptions, the only parameter that characterizes the source location is the DOA [102].

Suppose there are K signals present at an array of M sensors, $K < M$. The DOAs of the signals are $\theta_1, \theta_2, \dots, \theta_K$. The sensor outputs are appropriately pre-processed and sampled at arbitrary time instants n , $n = 1, \dots, N$. Based on the simplifying assumptions above, the array output vector $\mathbf{x}(n)$, also called *a snapshot*, may be modeled as [102, 75]

$$\mathbf{x}(n) = A(\boldsymbol{\theta})\mathbf{s}(n) + \mathbf{v}(n). \quad (2.1)$$

Here $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_K]^T$, $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_K(n)]^T$ is the K -vector of signal waveforms, $A(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_K)]$ is an $M \times K$ matrix of steering vectors related to the DOAs and $\mathbf{v}(n)$ is an $M \times 1$ noise vector. The noise is assumed to be independent of the signals, zero mean and spatially and temporally white, with variance σ^2 . The signal vector $\mathbf{s}(n)$ is modeled as either stochastic or deterministic, depending on the application. For notational convenience we simply write A instead of $A(\boldsymbol{\theta})$ when there is no possibility of confusion. If $\hat{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$, then we also write \hat{A} instead of $A(\hat{\boldsymbol{\theta}})$.

The vector $\mathbf{a}(\theta)$ is given as

$$\mathbf{a}(\theta) = [g_1(\theta)e^{-j\omega_c\tau_1(\theta)}, g_2(\theta)e^{-j\omega_c\tau_2(\theta)}, \dots, g_M(\theta)e^{-j\omega_c\tau_M(\theta)}]^T$$

where ω_c is the center frequency, $g_k(\theta)$ represents the sensitivity of the k th sensor to the DOA θ and $\tau_k(\theta)$ is the time delay of the signal coming from DOA θ at the k th sensor relative to some reference point. We assume that the transfer characteristics and positions of the sensors are known, i.e. $\mathbf{a}(\theta)$ is function of θ only, as indicated by notation.

The collection of steering vectors over the parameter space of interest, Θ , is the *array manifold*, \mathcal{A} ,

$$\mathcal{A} = \{\mathbf{a}(\theta) \mid \theta \in \Theta\}.$$

It is assumed that the mapping from Θ to the steering vectors is one-to-one and for $K + 1$ distinct DOAs $\theta_1, \dots, \theta_K, \theta_{K+1}$ the corresponding steering vectors are linearly independent.

The functions $\{g_k(\theta)\}_{k=1}^M$ depend on the type of sensors being used. If the sensors are omnidirectional, the sensors have equal sensitivity to all directions, implying that $\{g_k(\theta)\}_{k=1}^M$ are independent of θ . Sometimes the sensors may also be considered identical. By taking the first sensor as a reference element and redefining the signal vector in an appropriate manner ($g_k(\theta_k)s_k(n)$ is redefined as $s_k(n)$, $k = 1, \dots, K$), we can write

$$\mathbf{a}(\theta) = [1, e^{-j\omega_c\tau_2(\theta)}, \dots, e^{-j\omega_c\tau_M(\theta)}]^T.$$

We now consider the array of M identical sensors uniformly spaced on a line. Such an array is commonly referred to as a Uniform Linear Array (ULA). The array scheme is presented in Figure 2.1. The DOAs are measured counterclockwise from the line joining the sensors. In this case $\Theta = [0, \pi]$ and $\mathbf{a}(\theta)$ is given by [102]

$$\mathbf{a}(\theta) = [1, e^{2\pi j(d/\lambda)\cos(\theta)}, \dots, e^{2\pi j(M-1)(d/\lambda)\cos(\theta)}]^T$$

where d denotes the element spacing and λ denotes the wavelength. Note that we have to assume $d \leq \lambda/2$ to make $\mathbf{a}(\theta)$ uniquely defined (i.e., to avoid “spatial aliasing”). The restriction $\Theta = [0, \pi]$ is needed because two sources at locations symmetric with respect to the array line yield identical set of delays and hence cannot be distinguished from one another. In practice this ambiguity of ULAs is eliminated by using sensors that only pass signals whose DOAs are in $[0, \pi]$ [102].

If the assumptions made so far cannot be satisfied, significant changes in the signal

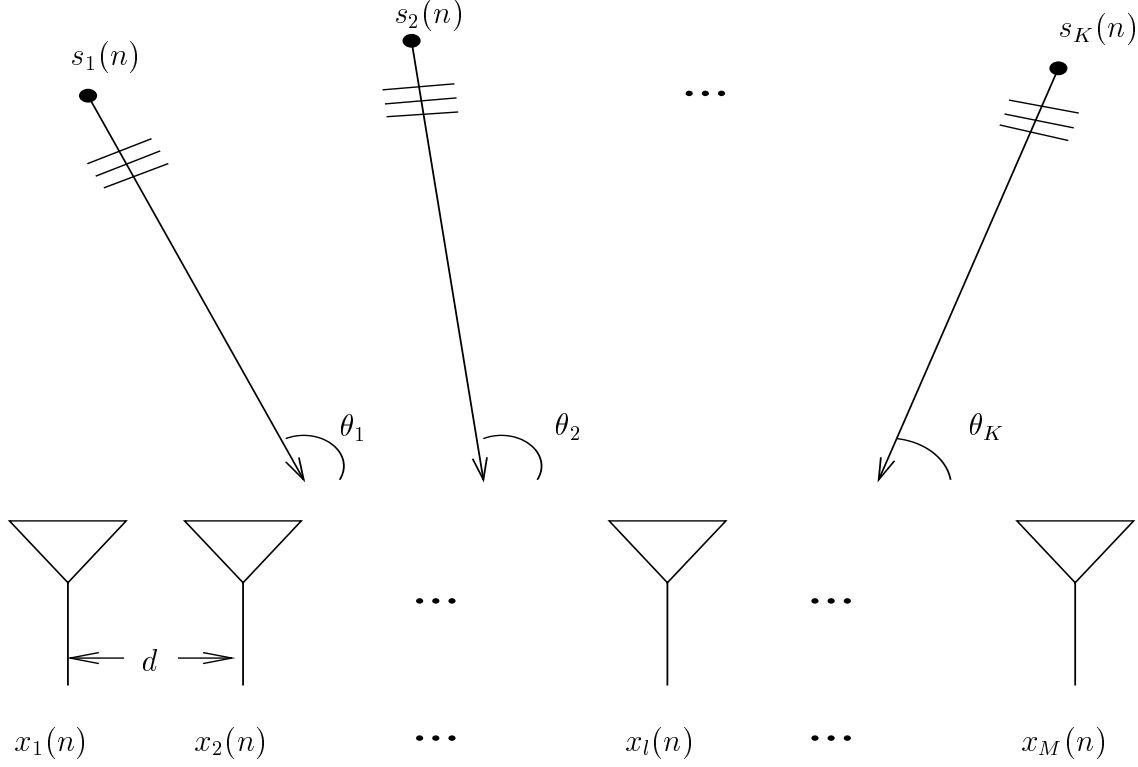


Figure 2.1: A uniform linear array of M sensors receiving plane waves from K far-field point sources.

model result. For example, if the signals are wide-band or located at the near field, the derived signal model is generally not valid.

2.2 Review of DOA estimation methods

In this section the most well-known DOA estimation techniques are addressed. We start by introducing the conventional beamformer and Capon's minimum variance method. The classical subspace-methods MUSIC and ESPRIT and their modifications are then reviewed. An important special case of estimating DOAs of coherent signals is considered as well. The remainder of the section considers maximum likelihood and subspace fitting methods. The performance of the reviewed techniques is considered in terms of resolution, large sample properties and computational complexity.

2.2.1 The Conventional beamformer

The conventional beamformer is one of the older techniques for localizing signal sources. The idea is to “steer” the array in one direction at a time and measure the output power. The steering directions which result in maximum power at the output provide the DOA estimates [55]. Steering is done by forming a linear combination of the sensor outputs

$$y(n) = \mathbf{w}^H \mathbf{x}(n).$$

Suppose there is only one wide sense stationary signal present in the array’s field of view and the signal is arriving from direction θ_1 . The optimal beamforming weight vector \mathbf{w}_{BF} is derived by maximizing the power of the output $y(n)$, subject to $\mathbf{w}^H \mathbf{w} = 1$. Using the signal model (2.1), the array output vector is given as

$$\mathbf{x}(n) = \mathbf{a}(\theta_1)s_1(n) + \mathbf{v}(n).$$

Recalling that the noise was assumed to be spatially and temporally white and independent of the signals, the power of $y(n)$ is given as

$$E\{y(n)y^*(n)\} = \mathbf{w}^H \Sigma \mathbf{w} = \sigma_s^2 \mathbf{w}^H \mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1) \mathbf{w} + \sigma_v^2 \mathbf{w}^H \mathbf{w}, \quad (2.2)$$

where $\Sigma = E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ is the array output covariance matrix, $\sigma_s^2 = E\{s_1(n)s_1^*(n)\}$ is the signal power and σ^2 is the noise variance. The problem of maximizing the output power may now be formulated as

$$\max_{\mathbf{w}} \{ \mathbf{w}^H \mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1) \mathbf{w} \} = \max_{\mathbf{w}} |\mathbf{w}^H \mathbf{a}(\theta_1)|^2, \quad \text{subject to } \mathbf{w}^H \mathbf{w} = 1,$$

where $|\cdot|$ is the modulus. The Cauchy-Schwarz inequality [102, p. 273] and the condition $\mathbf{w}^H \mathbf{w} = 1$ then imply that

$$|\mathbf{w}^H \mathbf{a}(\theta_1)|^2 \leq \|\mathbf{w}\|^2 \|\mathbf{a}(\theta_1)\|^2 = \|\mathbf{a}(\theta)\|^2,$$

where $\|\cdot\|$ is the Euclidean vector norm. The equality is obtained by choosing

$$\mathbf{w}_{BF} = \frac{\mathbf{a}(\theta_1)}{\sqrt{\mathbf{a}^H(\theta_1) \mathbf{a}(\theta_1)}}.$$

Hence \mathbf{w}_{BF} is the optimal weighting vector.

Inserting the optimal weighting vector in (2.2), we obtain the spatial spectrum

$$V_{BF}(\theta_1) = \frac{\mathbf{a}^H(\theta_1)\Sigma\mathbf{a}(\theta_1)}{\mathbf{a}^H(\theta_1)\mathbf{a}(\theta_1)}. \quad (2.3)$$

Note that among the all possible DOAs the correct DOA θ_1 gives the maximum of the above expression. Therefore when θ_1 is not known, it can be found by maximizing the above expression with respect to θ .

In practice, the covariance matrix Σ has to be estimated from the observed data $\mathbf{x}(1), \dots, \mathbf{x}(N)$ and the usual estimate of Σ is the sample covariance matrix

$$S = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^H(n).$$

Finally, the DOA estimate is chosen to be the angle of the highest peak in the estimated spatial spectrum

$$\hat{V}_{BF}(\theta) = \frac{\mathbf{a}^H(\theta)S\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{a}(\theta)}. \quad (2.4)$$

The estimate obtained is consistent.

When $K > 1$ signals are present $V_{BF}(\theta)$ should give a good indication of the energy coming from the direction θ . Hence $V_{BF}(\theta)$ should peak at the correct DOAs. In this general case the conventional beamforming DOA estimates are chosen to be the angles of the K highest peaks in (2.4). Naturally the estimates obtained are not consistent.

The conventional beamformer can not resolve two signals with close angles of arrival regardless of the available data quality or amount, i.e. its *resolution* is limited. It can be shown that for a ULA of M sensors, the beamforming resolution limit is approximately $\frac{\lambda}{Md}$ [102]. For example, for a ULA of 6 sensors of half-wavelength inter-element spacing, the approximate resolution limit equals $1/3$ rad $\approx 19^\circ$. Note that the low resolution also limits the number of DOAs that can be estimated.

2.2.2 Capon's method

Capon's minimum variance method [15] is a beamforming technique that attempts to overcome the poor resolution problems associated with the conventional beamformer. Intuitive idea of this method is based on the notion that when multiple sources are present, the

power measured by the spatial spectrum (2.3) is not only due to the power of the source at that direction, but also to power of other sources in other directions. This property limits the resolution of the conventional beamformer. Capon's method attempts to overcome the contribution of the undesired interferences by minimizing the total output power of $y(n) = \mathbf{w}^H \mathbf{x}(n)$ while maintaining a constant gain in the look direction θ .

Assume that $\mathbf{x}(n)$ given in (2.1) is a wide sense stationary random process and let $E\{\mathbf{x}(n)\mathbf{x}^H(n)\} = \Sigma$. The constrained minimization problem can be given mathematically as follows

$$\min_{\mathbf{w}} \mathbf{w}^H \Sigma \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{a}(\theta) = 1. \quad (2.5)$$

Using elementary analysis and linear algebra [102, p. 283], the solution to (2.5) is

$$\mathbf{w}_{CAP} = \frac{\Sigma^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \Sigma^{-1} \mathbf{a}(\theta)}.$$

The weights obtained are often called the Minimum Variance Distortionless Response (MVDR) beamformer weights. In practice, an estimate for \mathbf{w}_{CAP} is formed from the snapshots $\mathbf{x}(1), \dots, \mathbf{x}(N)$ by

$$\hat{\mathbf{w}}_{CAP} = \frac{S^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) S^{-1} \mathbf{a}(\theta)}.$$

Using the weights $\hat{\mathbf{w}}_{CAP}$, the DOA estimates obtained from the Capon's method are chosen to be the K highest peaks in the spectrum

$$\hat{V}_{CAP}(\theta) = \frac{1}{\mathbf{a}(\theta)^H S^{-1} \mathbf{a}(\theta)}. \quad (2.6)$$

Capon's method has been found empirically to possess superior performance to that of the conventional beamformer. See Fig. 2.2 for comparison of the conventional beamformer and Capon's method in the situation where two independent random 4-QAM signals of equal power (SNR is 20 dB) from directions 81° and 99° arrive to a 6-element ULA with interelement spacing equal to half a wavelength. In this example the number of snapshots is $N = 300$ and the marginals of the noise are i.i.d. circular complex Gaussian.

The resolution of the Capon's method is, however, still dependent upon the number of sensors and on the SNR. It has also been shown that performance of the method severely

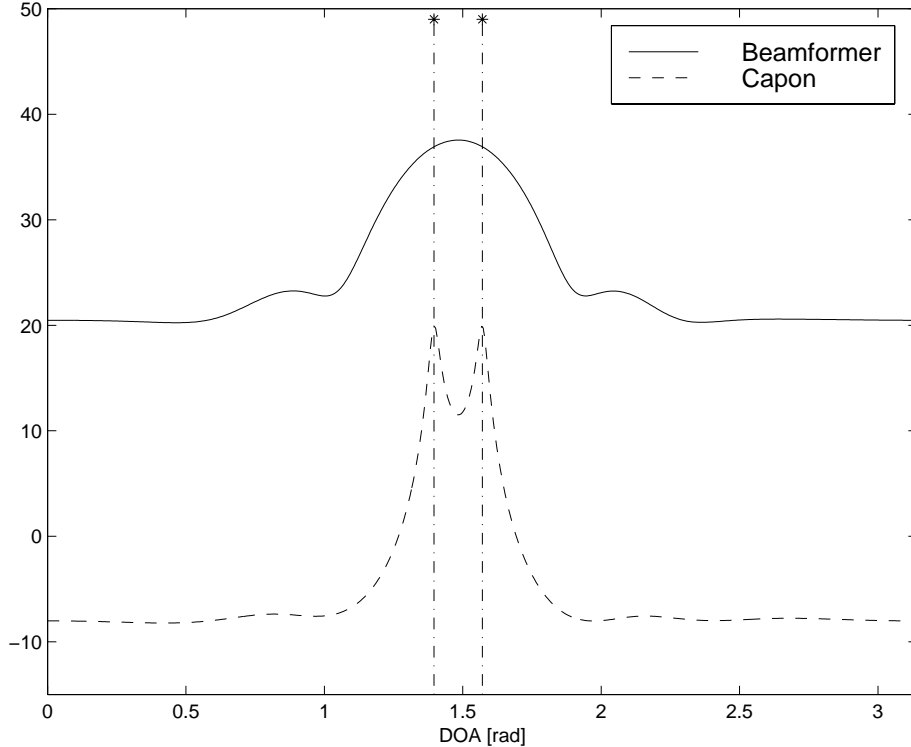


Figure 2.2: Comparison of the conventional beamformer and Capon's method. Two independent random 4-QAM signals of equal power (SNR is 20 dB) from directions 81° and 99° arrive at a 6-element ULA with interelement spacing equal to half a wavelength.

degrades in the case of correlated signals because Capon's method couples the correlated signals ineffectively to reduce the output variance [52]. Note that the conventional beamformer and Capon's method can be derived without using the parametric signal model (2.1) as was done in [102]. Therefore these methods are sometimes called as nonparametric DOA estimation methods.

2.2.3 Subspace methods

So-called *subspace* DOA estimation methods have been the focus of much research since Schmidt [96] introduced the MUSIC algorithm in 1979. The reason for their popularity is that the subspace methods can, in theory, resolve sources with arbitrarily close DOAs. Prior to introducing these methods, we discuss some basic assumptions and terminology.

Consider the signal model given in (2.1). Assume that the K -dimensional signal vec-

tor $\mathbf{s}(n)$ is a wide sense stationary process and the $K \times K$ signal covariance matrix $\Sigma_s = E\{\mathbf{s}(n)\mathbf{s}^H(n)\}$ is of rank K , i.e. the signals are *incoherent*. The signals are incoherent unless one of them is a scaled version of the other i.e. their correlation coefficient $\rho = \pm 1$. When $\rho = \pm 1$ the signals are said to be *coherent*. Alternatively one could consider deterministic signals and assume that the limiting sample signal covariance matrix $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbf{s}(i)\mathbf{s}^H(i)$ is a constant matrix of rank K . In this section we, however, model the signals as stochastic and wide sense stationary. The covariance matrix of $\mathbf{x}(n)$ is

$$\Sigma = E[\mathbf{x}(n)\mathbf{x}^H(n)] = A\Sigma_s A^H + \sigma^2 I \quad (2.7)$$

where σ^2 is the noise variance. Consequently the $M - K$ smallest eigenvalues of Σ are equal to σ^2 and the corresponding eigenvectors are orthogonal to the columns of A . These eigenvectors span the *noise subspace* and the eigenvectors corresponding to the K largest eigenvalues span the *signal subspace*.

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K > \lambda_{K+1} = \dots = \lambda_M = \sigma^2$ be the eigenvalues of Σ and let $\mathbf{u}_1, \dots, \mathbf{u}_M$ be the corresponding eigenvectors. Denote the $M \times K$ matrix of the signal subspace eigenvectors by $U_s = [\mathbf{u}_1, \dots, \mathbf{u}_K]$ and the $M \times (M - K)$ matrix of the noise subspace eigenvectors by $U_n = [\mathbf{u}_{K+1}, \dots, \mathbf{u}_M]$. The projection matrix to the signal subspace is $\Pi_A = U_s(U_s^H U_s)^{-1} U_s^H = U_s U_s^H$. Because the columns of A also span the signal subspace, $\Pi_A = A A^\dagger$, where $A^\dagger = (A^H A)^{-1} A^H$. The projection matrix to the noise subspace is given by $\Pi_A^\perp = U_n U_n^H = I - \Pi_A$. Note that the covariance matrix Σ may also be given as

$$\Sigma = [U_s \ U_n] \text{diag}\{\lambda_1, \dots, \lambda_M\} [U_s \ U_n]^H,$$

where

$$\text{diag}\{\lambda_1, \dots, \lambda_M\} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_M \end{bmatrix}.$$

MUSIC

The MUSIC (Multiple Signal Classification) algorithm proposed by Schmidt [96] is a noise subspace algorithm. The technique is an extension of the Pisarenko harmonic decomposition for estimating the frequencies of a sum of complex exponentials in white noise [80]. It exploits the orthogonality of the noise subspace eigenvectors and the columns of the matrix A .

Because of the orthogonality of the signal and noise subspace,

$$\mathbf{a}^H(\theta_i)U_n U_n^H \mathbf{a}(\theta_i) = 0$$

at the DOAs θ_i , $i = 1, \dots, K$. Furthermore, by using the assumption that the steering vectors corresponding to $K + 1$ different DOAs are linearly independent, it is easy to show that the above relation holds only at these points (see [102, p. 157]). When U_n is estimated using the matrix of the eigenvectors corresponding to the $M - K$ smallest eigenvalues of the sample covariance matrix S , say \hat{U}_n , the pseudo-spectrum

$$\hat{V}_M(\theta) = \frac{1}{\mathbf{a}^H(\theta)\hat{U}_n \hat{U}_n^H \mathbf{a}(\theta)}$$

will exhibit large peaks at the correct DOAs due to the orthogonality. In the MUSIC algorithm, the estimates of the DOAs are chosen to be the K largest peaks in this pseudo-spectrum.

MUSIC can, in theory, resolve sources with arbitrary close DOAs. The maximum number of DOAs that can be estimated with an M element array is $M - 1$. In contrast to the beamforming techniques, the MUSIC algorithm provides *consistent* estimates [103]. Figure 2.3 represents a comparison of Capon's method and MUSIC in the situation where two independent random 4-QAM signals of equal power (SNR is 20 dB) from directions 88° and 92° arrive to a 6-element ULA with interelement spacing equal to half a wavelength. As in the previous example the number of snapshots is $N = 300$ and the marginals of the noise are i.i.d. circular complex Gaussian.

Statistical properties of the MUSIC estimation technique have been widely studied. Stoica and Nehorai [103] derived the asymptotic covariance matrix of the MUSIC estimates

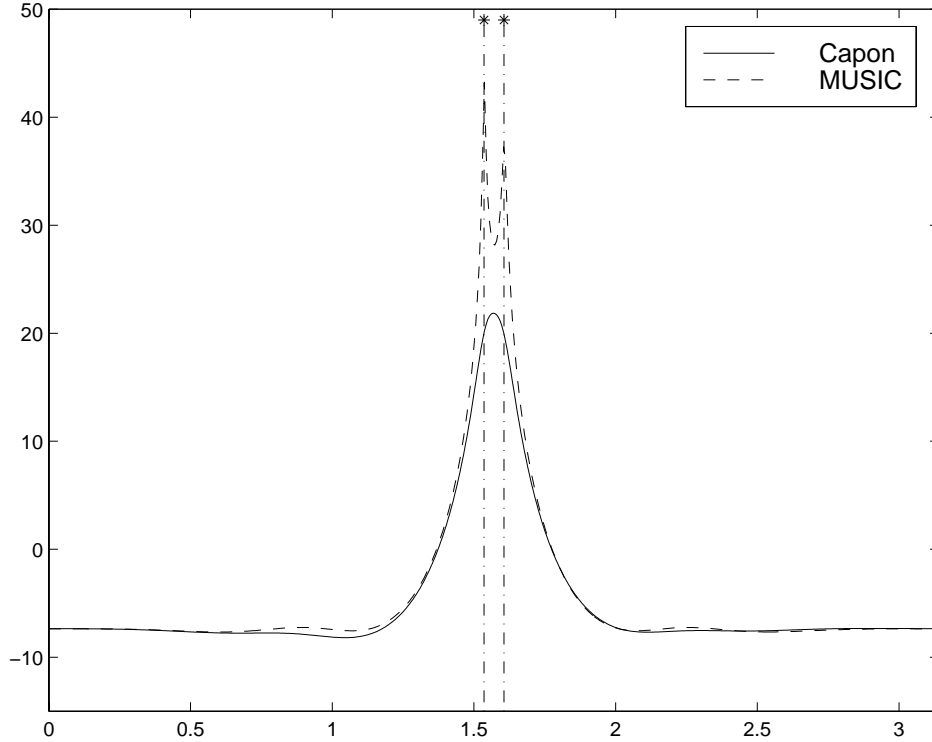


Figure 2.3: Comparison of Capon’s method and MUSIC. Two independent random 4-QAM signals of equal power (SNR is 20 dB) from directions 88° and 92° arrive to a 6-element ULA with interelement spacing equal to half a wavelength

assuming deterministic signals and Gaussian noise. They also derived the Cramér-Rao bound for DOA estimates and showed that the MUSIC estimator does not asymptotically achieve it for finite M . Their analysis also shows that the best asymptotic performance of MUSIC is obtained when the signals are uncorrelated. When the signals become correlated, the performance of the MUSIC estimator degrades. See also [103, 105, 48, 128, 125, 82].

There are plenty of modifications of MUSIC. So-called weighted MUSIC applies weights to the noise subspace eigenvectors used in the MUSIC algorithm. It includes as a special case the Min-Norm [57] method. Root-MUSIC algorithm [4] is based on polynomial rooting. It reduces the computational requirements of MUSIC and provides higher resolution than MUSIC for small sample sizes, but is applicable only to uniform linear arrays. An approach where MUSIC algorithm is applied to spatially prefiltered observations (beam-space MUSIC) is presented in [10].

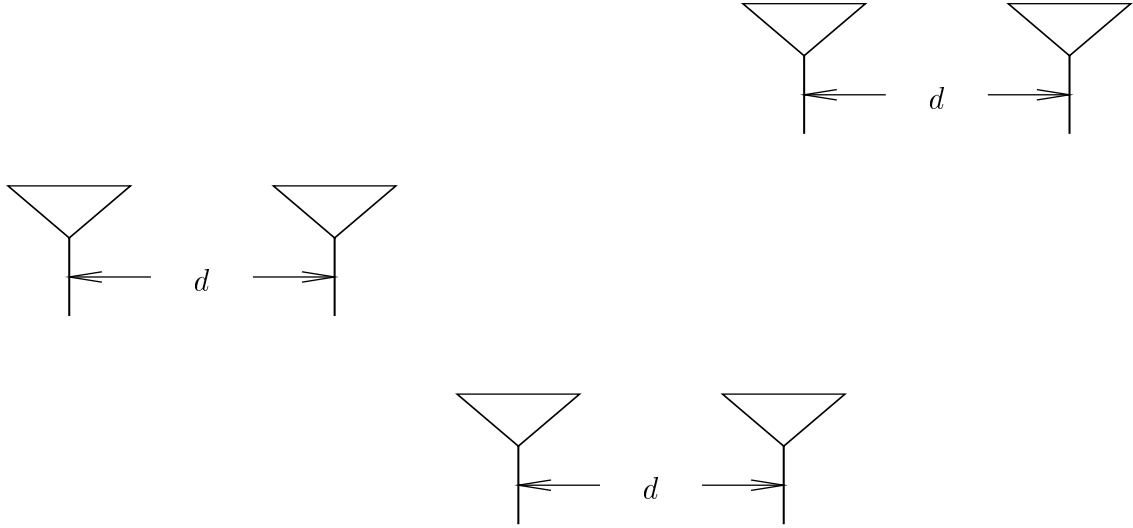


Figure 2.4: Illustration of ESPRIT array geometry: three pairs of sensors forming two subarrays.

ESPRIT

ESPRIT (Estimation of Signal Parameters via Rotation Invariant Techniques) [76, 94] is a signal subspace technique. It dramatically reduces the computational and storage requirements of MUSIC and does not involve an exhaustive search through all possible steering vectors to estimate the DOAs. Moreover, ESPRIT does not require that the matrix A is precisely known.

ESPRIT requires that the sensor array can be decomposed into two identical subarrays separated by some fixed displacement vector. The subarrays can also overlap. See Fig. 2.4 for illustration of required array geometry.

Consider again the signal model (2.1). Denote the dimension of the twin subarrays by P . Let J_1 be the $P \times M$ matrix that selects the leftmost subarray from the array output vector and denote the corresponding matrix for rightmost subarray by J_2 . In the case of ULAs, for instance, it is common to form the first subarray from the sensors labeled $1, \dots, M-1$ and the second subarray from the sensors labeled $2, \dots, M$, so that the selection matrices J_1 and J_2 are given as $J_1 = [I_{M-1} \mathbf{0}]$ and $J_2 = [\mathbf{0} I_{M-1}]$.

The basis of the ESPRIT algorithm is the observation that

$$J_2 A = J_1 A \Phi, \quad (2.8)$$

where Φ is a diagonal matrix with diagonal elements ϕ_i given by

$$\phi_i = \exp\{j2\pi(d/\lambda)\cos(\theta_i)\}, \quad i = 1, \dots, K. \quad (2.9)$$

Note that Φ is a unitary matrix. By assuming that $d \leq \lambda/2$ and $\Theta = [0, 2\pi]$, the DOA θ_i is related to ϕ_i by

$$\theta_i = \arccos(-j \frac{\lambda}{2\pi d} \arg\{\phi_i\}). \quad (2.10)$$

Let U_s be the matrix of signal subspace eigenvectors. Because U_s and A span the same column space, there has to be a unique non-singular $K \times K$ matrix C such that $U_s = AC$.

We now form a set of equations

$$\begin{cases} J_1 U_s = J_1 A C \\ J_2 U_s = J_2 A C. \end{cases} \quad (2.11)$$

Using (2.8) in (2.11), the following relation is obtained

$$J_2 U_s = J_1 U_s \Psi, \quad (2.12)$$

where Ψ is given by $\Psi = C^{-1} \Phi C$. A standard result from matrix algebra states that Ψ and Φ have the same eigenvalues [37, p. 525] (matrices are similar), i.e. the eigenvalues of Ψ are equal to the diagonal elements of Φ . In theory, the DOAs can therefore be solved by finding Ψ such that the relation (2.12) is true and then applying inverse mapping (2.10) to the eigenvalues of Ψ .

In practice, an estimate \hat{U}_s for U_s has to be formed from finite number of noisy observations and finding Ψ such that

$$J_2 \hat{U}_s = J_1 \hat{U}_s \Psi$$

is not possible. Therefore some approximation method must be used. First solution was to use the Least Squares method and the resulting algorithm is called LS-ESPRIT. Because both sides of the above equation contain similar “error”, the Total Least Squares (TLS) method suits better for this purpose. An algorithm where TLS method is used is called

TLS-ESPRIT. TLS-ESPRIT method needs more computation than LS-ESPRIT because the TLS solution requires calculation of singular value decomposition.

The number of DOAs that can be estimated depends on the array structure. By using overlapping subarrays of size $M - 1$, it is possible to estimate up to $M - 1$ DOAs. Swindlehurst et al. [108] discuss the optimal choice of the two subarrays when the choice is not unique. Similarly to MUSIC, the estimates obtained using ESPRIT are consistent (under some general assumptions) and the resolution of ESPRIT is not limited. Stoica and Nehorai [106] derived the asymptotic covariance matrix of the ESPRIT estimates (under Gaussian signal and noise assumption). They also showed that the asymptotic variance of the ESPRIT estimates is always larger than the asymptotic variance of the MUSIC estimators and the difference is notable for large arrays and small number of signals. See also [83, 73, 127].

The ESPRIT technique is theoretically equivalent to the Toeplitz Approximation Method [3, 58]. The Unitary ESPRIT algorithm [33] is based on transformation of the complex data matrix to the real matrix of the same size and involves only real calculations. ESPRIT-type algorithms based on fourth order cumulants are discussed in [18, 26].

Coherent signals

If some of the signal sources are completely coherent, i.e. their correlation is equal to ± 1 , the signal covariance matrix is singular and the conventional subspace methods do not work anymore. Coherent signals can arise from multipath propagation where, due to reflections, the same signal arrives to an array from multiple directions. Smart jammers can also create coherent signals.

In the case of ULA, the DOAs can, however, be estimated by using spatial smoothing preprocessing. In the following we briefly describe this technique. For simplicity, we will drop the time index from \mathbf{x} , \mathbf{s} and \mathbf{v} from this point onward.

The idea in the spatial smoothing scheme is to divide a linear uniform array with M identical sensors into overlapping forward and backward subarrays of size P , $M \geq P > K$.

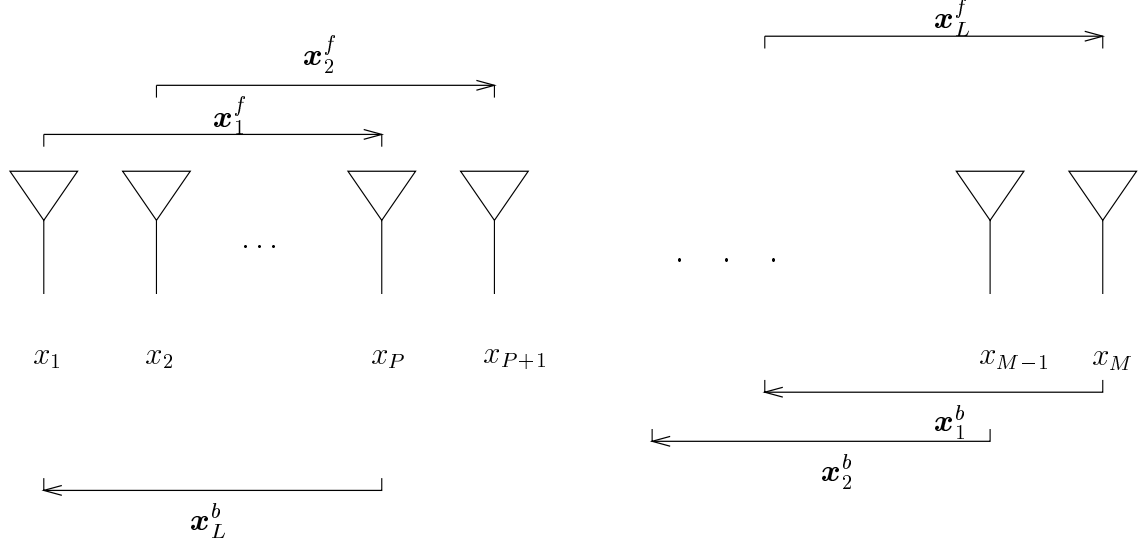


Figure 2.5: The forward/backward spatial smoothing scheme.

See Figure 2.5, [77, 78]. Let \mathbf{x}_l^f denote the received signals at the l th forward subarray i.e.

$$\mathbf{x}_l^f = (x_l, \dots, x_{l+P-1})^T. \quad (2.13)$$

Moreover, let \mathbf{x}_l^b denote the complex conjugate of the received signals at the l th backward subarray

$$\mathbf{x}_l^b = (x_{M-l+1}^*, \dots, x_{M-P-l+2}^*)^T, \quad (2.14)$$

and let D^l denote the l th power of the diagonal matrix

$$D = \text{diag}\{e^{-j2\pi(d/\lambda)\cos(\theta_1)}, \dots, e^{-j2\pi(d/\lambda)\cos(\theta_K)}\}. \quad (2.15)$$

Adapting the same notation as before, we can model \mathbf{x}_l^f as

$$\mathbf{x}_l^f = AD^{(l-1)}\mathbf{s} + \mathbf{v}_l^f$$

and

$$\mathbf{x}_l^b = AD^{(l-1)}(D^{(M-1)}\mathbf{s})^* + \mathbf{v}_l^b$$

where $A = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ with $\mathbf{a}(\theta_k)$ being the $P \times 1$ ($P > K$) array steering vector corresponding to the DOA of the k th signal component, and \mathbf{v}_l^f and \mathbf{v}_l^b are noise vectors.

The forward-averaged spatially-smoothed covariance matrix Σ^f is defined as the average

of the subarray covariance matrices

$$\Sigma^f = \frac{1}{L} \sum_{l=1}^L \Sigma_l^f,$$

where $\Sigma_l^f = E\{\mathbf{x}_l^f \mathbf{x}_l^{fH}\}$ and $L = M - P + 1$. In addition, the backward averaged spatially smoothed covariance matrix is

$$\Sigma^b = \frac{1}{L} \sum_{l=1}^L \Sigma_l^b,$$

where $\Sigma_l^b = E\{\mathbf{x}_l^b \mathbf{x}_l^{bH}\}$. Finally, the forward/backward spatial smoothed covariance matrix $\bar{\Sigma}$ is defined as

$$\bar{\Sigma} = \frac{\Sigma^f + \Sigma^b}{2}. \quad (2.16)$$

Assume $M \geq 3K/2$. It has been shown that under mild restrictions on \mathbf{s} , it is possible to choose P such that the $P - K$ smallest eigenvalues of $\bar{\Sigma}$ are equal and the corresponding eigenvectors are orthogonal to the columns of the matrix A [77]. Therefore the DOAs of the coherent signals can be estimated using any subspace algorithm and an estimate of $\bar{\Sigma}$. If $L \geq K$, the restriction on \mathbf{s} is not needed.

The number of DOAs that can be estimated with a ULA of M sensors and using forward/backward spatial smoothing preprocessing is $\lfloor 2M/3 \rfloor$, under some general assumptions. The effect of forward/backward spatial smoothing on the performance of subspace methods has been studied in [79, 86, 87, 84, 85, 109].

The forward only spatial smoothing is a special case of a preprocessing technique called Autocorrelation Matrix Smoothing (AMS). The AMS can be viewed as a general framework for techniques reducing the cross-correlations between the arriving signals. See [53] and references therein.

2.2.4 Maximum likelihood methods

Probably the most popular technique for parameter estimation is the Maximum Likelihood (ML) method. In this section we consider two ML techniques for DOA estimation which are derived under Gaussian assumption on the array output vector. The two techniques are Stochastic Maximum Likelihood (SML) and Deterministic Maximum Likelihood (DML).

The SML technique models the signal and noise as Gaussian and the DML method is based on deterministic signal and Gaussian noise assumptions. In the literature, also the terms “unconditional ML” and “conditional ML” are used for SML and DML respectively [104].

In the SML approach, the signal vector is modeled as a zero mean and temporally white circular Gaussian random process, i.e. its second order moments are given by

$$E\{\mathbf{s}(l)\mathbf{s}^H(k)\} = \Sigma_s \delta_{lk} \quad \text{and} \quad E\{\mathbf{s}(l)\mathbf{s}^T(k)\} = 0$$

where δ_{lk} is the Kronecker delta. This condition implies that the real and imaginary parts of any marginal component of $\mathbf{s}(n)$ are independent and identically distributed. The rank of Σ_s is $K' \leq K$. It is assumed that $K < (M + K')/2$ to ensure parameter identifiability (see [75, 120, 72]). The noise $\mathbf{v}(n)$ is assumed to be zero mean spatially and temporally white circular Gaussian process, i.e.

$$E\{\mathbf{v}(l)\mathbf{v}^H(k)\} = \sigma^2 I \delta_{lk} \quad \text{and} \quad E\{\mathbf{v}(l)\mathbf{v}^T(k)\} = 0.$$

Moreover, the signal and noise are assumed to be mutually independent. The Gaussian assumption implies that the negative log likelihood function of the snapshot data $\mathbf{x}(1), \dots, \mathbf{x}(n)$ is proportional to [75]

$$D(\boldsymbol{\theta}, \Sigma_s, \sigma^2) = \log[\det\{\Sigma\}] + \text{Tr}\{\Sigma^{-1}S\} = \log[\det\{A\Sigma_s A^H + \sigma^2 I\}] + \text{Tr}\{(A\Sigma_s A^H + \sigma^2 I)^{-1}S\} \quad (2.17)$$

where \det stands for the determinant, Tr denotes the trace, and $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^T$. This criterion allows explicit separation of some of the parameters. For fixed $\boldsymbol{\theta}$, the minimum with respect to σ^2 and Σ_s can be shown to be [6, 8, 44]

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{M - K} \text{Tr}\{\Pi_A^\perp S\} \quad (2.18)$$

$$\hat{\Sigma}_s(\boldsymbol{\theta}) = A^\dagger (S - \hat{\sigma}^2(\boldsymbol{\theta})I) A^{\dagger H}. \quad (2.19)$$

By substituting these estimates into (2.17), the stochastic maximum likelihood estimate of $\boldsymbol{\theta}$ is of the form

$$\hat{\boldsymbol{\theta}}_{SML} = \arg\{\min_{\boldsymbol{\theta}} \log[\det\{A\hat{\Sigma}_s(\boldsymbol{\theta})A^H + \hat{\sigma}^2(\boldsymbol{\theta})I\}]\}. \quad (2.20)$$

The estimate for the signal covariance matrix is not guaranteed to be positive semidefinite [74]. This is because the equations (2.18)-(2.20) are derived considering the minimization of the criterion function (2.17) with respect to Σ_s over the set of Hermitian matrices, and not over the set of Hermitian positive semidefinite matrices. The estimates obtained from (2.18)-(2.20) are correctly termed as a *large-sample realizations* of the ML estimates of $\boldsymbol{\theta}$, Σ_s and σ^2 . For further details, see [107].

The DML technique models the emitter signals as unknown deterministic waveforms and the noise as a zero mean spatially and temporally white circular Gaussian random process. The deterministic ML estimate for the DOAs is of the form

$$\hat{\boldsymbol{\theta}} = \arg\{\min_{\boldsymbol{\theta}} \text{Tr}\{\Pi_A^\perp S\}\}. \quad (2.21)$$

In the one source signal case the DML technique is equivalent to the conventional beam-former if $\|\mathbf{a}(\theta)\|$ is independent of θ . After finding the estimate $\boldsymbol{\theta}$, the ML estimates for the signal waveforms and noise variance can be calculated by

$$\hat{\sigma}^2 = \frac{1}{M-K} \text{Tr}\{\Pi_A^\perp S\}, \quad \hat{\mathbf{s}}(n) = A^\dagger \mathbf{x}(n).$$

Note that the SML and DML can also deal with coherent signals. The large sample properties of the two ML methods are well reported. It has been shown that under some “regularity conditions”, the SML DOA estimation method achieves the Cramer-Rao bound as the number of observations tends to infinity [74, 104]. This means that the SML method is asymptotically efficient. The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{SML}$ is unchanged if the signals are modeled as deterministic, i.e. the limiting performance of the SML method is the same for Gaussian and deterministic waveforms [74, 104]. The DML method does not achieve the Cramer-Rao bound when $N \rightarrow \infty$ for finite M [103]. Similar to the results for SML, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{DML}$ is the same for deterministic and Gaussian signals.

The criterion functions (2.17)-(2.21) are highly nonlinear functions of $\boldsymbol{\theta}$ and a multidimensional numerical search with sufficiently accurate initial estimates has to be applied. This makes the computational cost much higher than for the techniques introduced earlier in this chapter. The optimization techniques applied include the Expectation Maximization

(EM) algorithm [31, 70], alternative projection [132], the Iterative quadratic ML (IQML) method [9, 41] and Newton-type techniques [7, 12, 100, 75].

The DOA estimates obtained from ML techniques are usually more accurate than the estimates obtained from beamforming or subspace methods, especially for scenarios involving highly correlated or even coherent signals. For example, the MUSIC and DML are asymptotically equivalent if the signals are uncorrelated but for correlated signals the DML usually performs better [103, 106].

2.2.5 Subspace fitting methods

As discussed earlier, the SML and DML DOA estimates can be more accurate than the estimates obtained from subspace methods (under Gaussian assumption). On the other hand the ML methods require a multidimensional search and therefore a high computational cost. The *weighted subspace fitting* method has the same asymptotic behavior as the SML under the Gaussian assumption. Moreover, the computational load needed is less than for the SML. In this subsection we review some basic ideas of a general subspace fitting principle and discuss the asymptotic properties of resulting techniques.

Assume that the signal and noise vectors are distributed as given in the case of the SML method. Consider the array output covariance matrix given in (2.7)

$$\Sigma = A\Sigma_s A^H + \sigma^2 I.$$

Let the rank of the signal covariance matrix Σ_s be $K' \leq K$ and denote the eigenvalues and the corresponding eigenvectors of Σ by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{K'} > \lambda_{K'+1} = \dots = \lambda_M = \sigma^2$ and $\mathbf{u}_1, \dots, \mathbf{u}_M$, respectively. Assume $K < (M + K')/2$ as in the case of the SML method. Introduce the notation

$$\Lambda_s = \text{diag}[\lambda_1, \dots, \lambda_{K'}], \quad U_s = [\mathbf{u}_1, \dots, \mathbf{u}_{K'}], \quad U_n = [\mathbf{u}_{K'+1}, \dots, \mathbf{u}_M]$$

Recall that if the signal covariance matrix Σ_s is of full rank, i.e. $K' = K$, A and U_s span the same column space. If $K' < K$, the K' -dimensional column space of U_s is contained in the K -dimensional column space of A . This then implies that for $K' \leq K$, there exists a

unique $K \times K'$ matrix T such that

$$U_s = AT.$$

Using this relation, the signal subspace fitting criterion is given by

$$[\hat{\boldsymbol{\theta}}, \hat{T}] = \arg \min_{\boldsymbol{\theta}, T} \|\hat{U}_s - AT\|_W^2 \quad (2.22)$$

where $\|X\|_W^2 = \text{Tr}\{XWX^H\}$ and W is a $K' \times K'$ positive definite weighting matrix. The solution with respect to T is given by $\hat{T} = A^\dagger \hat{U}_s$ [75] and substituting it to (2.22) we get

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} D_{SSF}(\boldsymbol{\theta})$$

where

$$D_{SSF}(\boldsymbol{\theta}) = \|\hat{U}_s - AA^\dagger\|_W^2 = \text{Tr}\{\Pi_A^\perp \hat{U}_s W \hat{U}_s^H\}. \quad (2.23)$$

The above equations define the class of Signal Subspace Fitting (SSF) methods. Naturally, different choices of W lead to different estimates. The multidimensional MUSIC (MD-MUSIC) [12] estimates are obtained by choosing $W = I$. The optimal choice of W minimizing the estimation error variance can be shown to be [114]

$$W_{opt} = (\Lambda_s - \sigma^2 I)^2 \Lambda_s^{-1}.$$

Because W_{opt} depends on unknown parameters, an estimate of W_{opt} has to be used. This estimate is given by

$$\hat{W}_{opt} = (\hat{\Lambda}_s - \hat{\sigma}^2 I)^2 \hat{\Lambda}_s^{-1},$$

where $\hat{\sigma}^2$ is a consistent estimate of the noise variance, for example the average of the $M - K'$ smallest eigenvalues of S , and

$$\hat{\Lambda}_s = \text{diag}\{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{K'}\}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{K'}$ are the K' largest eigenvalues of S . The estimates obtained from minimizing (2.23) with weighting \hat{W}_{opt} are called Weighted Subspace Fitting (WSF) estimates. The WSF method has been shown to possess the same large sample properties as the SML method (for Gaussian signal and noise) at a lower computational cost provided a fast method for computing the eigendecomposition is used [55]. Also for ULAs a Root-WSF

algorithm [105] can be used to obtain a closed form solution. The limiting performance of WSF estimates is the same for Gaussian and deterministic signals [74].

Assume now that $K = K'$. Recall that in this case the columns of U_n are orthogonal to the columns of A . The Noise Subspace Fitting (NSF) criterion is obtained using this property, i.e.

$$U_n^H A = 0.$$

A natural estimate for $\boldsymbol{\theta}$ is obtained by minimizing the following criterion

$$D_{NSF}(\boldsymbol{\theta}) = \|\hat{U}_n^H A(\boldsymbol{\theta})\|_Q^2 = \text{Tr}\{Q A(\boldsymbol{\theta})^H \hat{U}_n \hat{U}_n^H A(\boldsymbol{\theta})\}$$

where Q is a $K \times K$ positive semidefinite weighting matrix. Different choices of Q lead to estimates with different asymptotic properties. If $Q = I$, the NSF method reduces to the MUSIC method. It can be shown [75] that the estimates obtained from the above noise subspace fitting criterion and the estimates obtained from SSF criterion are asymptotically equivalent for weights

$$Q = A^\dagger \hat{U}_s W \hat{U}_s^H A^{\dagger*}.$$

The NSF method can not deal with coherent signals and the optimal weighting matrix depends on $\boldsymbol{\theta}$. Therefore the WSF method is preferred.

2.3 Estimation of the number of signals

All the DOA estimation methods require the knowledge of the number of signals K . If this information is not provided *a priori*, it has to be estimated from the available data. The first techniques proposed for estimation of the number of signals were based on statistical hypotheses testing. Under the assumptions made in the case of the SML technique, the snapshots $\mathbf{x}(1), \dots, \mathbf{x}(N)$ are i.i.d. circular Gaussian with covariance matrix Σ . As before, let $\lambda_1 \geq \dots \geq \lambda_M$ denote the eigenvalues of Σ . The likelihood ratio test criterion for testing hypotheses

$$\mathcal{H}_0 : \quad \lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_M = \sigma^2$$

$$\mathcal{H}_1 : \quad \Sigma \text{ is arbitrary}$$

assuming σ^2 is unknown is [77]

$$L_1(k) = \frac{\frac{1}{M-k} \sum_{i=k+1}^M \hat{\lambda}_i}{\left(\prod_{i=k+1}^M \hat{\lambda}_i \right)^{1/(M-k)}},$$

where $\hat{\lambda}_i, i = 1, \dots, M$ are the eigenvalues of the sample covariance matrix S . Standard likelihood ratio testing theory states that, under \mathcal{H}_0 , $2N(M-k) \log L_1(k)$ is asymptotically χ^2 distributed with $(M-k)^2 - 1$ degrees of freedom [77]. The estimate of the number of signals can then be formed by calculating $2N(M-k) \log L_1(k)$ for $k = 0, 1, 2, \dots$ and choosing the estimate to be the first value for k that is smaller than some threshold value obtained from χ^2 distribution with appropriate degrees of freedom. Note that in the case of coherent signals the estimate obtained is the estimate of the rank of the signal covariance matrix K' .

Another possibility is to test between the hypotheses

$$\begin{aligned} \mathcal{H}_0 : \quad & \Sigma = A(\boldsymbol{\theta}) \Sigma_s A^H(\boldsymbol{\theta}) + \sigma^2 I, \quad \text{where } A \text{ is } M \times k \\ \mathcal{H}_1 : \quad & \Sigma \text{ is arbitrary} \end{aligned}$$

as discussed in [75]. The structured estimate for the covariance matrix under \mathcal{H}_0 is given by

$$\hat{\Sigma}(\hat{\boldsymbol{\theta}}) = A(\hat{\boldsymbol{\theta}}) \hat{\Sigma}_s(\hat{\boldsymbol{\theta}}) A^H(\hat{\boldsymbol{\theta}}) + \hat{\sigma}^2 I$$

where $\hat{\Sigma}_s$, $\hat{\sigma}^2$ and $\hat{\boldsymbol{\theta}}$ are the SML estimates obtained from equations (2.18)-(2.20). The likelihood ratio test criterion in this case is given by

$$L_2(k) = \frac{\det\{\hat{\Sigma}(\hat{\boldsymbol{\theta}})\}}{\det\{S\}}.$$

Under \mathcal{H}_0 , $2N \log(L_2(k))$ is asymptotically χ^2 distributed with $M^2 - k^2 - k - 1$ degrees of freedom [75]. The estimate of the number of signals can be obtained similarly as explained in the case of $L_1(k)$ (with remarkably higher computational cost due to the need for SML estimates). Note that the estimate obtained is always an estimate for the number of signals regardless of the coherence of the signals. The authors of [114] suggest a similar test procedure which is based on the WSF criterion.

A drawback for the estimation methods based on statistical hypothesis testing is the need for subjective threshold values. In the techniques proposed by Wax and Kailath [118] such threshold values are not needed. The proposed methods are based on information theoretic criteria, namely Akaike Information Criterion (AIC) [2] and Minimum Description Length (MDL) [88].

In the MDL based approach, the estimate of the number of signals is an integer $k \in \{0, 1, \dots, M-1\}$ which minimizes the criterion

$$MDL(k) = -\log \left(\frac{\left(\prod_{i=k+1}^M \hat{\lambda}_i \right)^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^M \hat{\lambda}_i} \right)^{(M-k)N} + \frac{1}{2}k(2M-k)\log N, \quad (2.24)$$

where $\hat{\lambda}_i, i = 1, \dots, M$ are the eigenvalues of the sample covariance matrix. In the case of AIC, the criterion function is

$$AIC(k) = -\log \left(\frac{\left(\prod_{i=k+1}^M \hat{\lambda}_i \right)^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^M \hat{\lambda}_i} \right)^{(M-k)N} + k(2M-k).$$

The methods were first derived by assuming that the signal and noise are mutually independent i.i.d. Gaussian processes [118]. The authors proved that, under this assumption, the MDL method gives a strongly consistent estimate of the number of signals. The AIC criterion was shown to overestimate the number of signals. Later Zhao et. al. [129] proved that strong consistency of the MDL method is obtained also when the assumption of i.i.d. Gaussian observations is violated. They also corrected the original consistency proof.

When forward/backward spatial smoothing is performed as a preprocessing step, the number of signals can be estimated by using a modified MDL criterion [124]. The estimate of the number of signals is an integer $k \in \{0, 1, \dots, M-1\}$ which minimizes the criterion

$$MDL_{f/b}(k) = -\log \left(\frac{\left(\prod_{i=k+1}^M \hat{\lambda}_i \right)^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^M \hat{\lambda}_i} \right)^{(M-k)N} + \frac{1}{4}k(2M-k+1)\log N, \quad (2.25)$$

where $\hat{\lambda}_i, i = 1, \dots, M$ are the eigenvalues of the forward/backward averaged sample covariance matrix.

Other types of methods appear in the literature. Wax and Ziskind [119] have used combined DOA and number of signals estimation based on the DML assumptions and the

MDL principle. A similar technique based on the SML assumption and MDL is presented in [116]. These methods can be more accurate than the MDL method but need multidimensional search for DOA estimation and thus have a high computational cost. An approach based on transformed Gerschgorin radii is presented in [123]. Also neural networks have been applied to number of signals estimation, see [20]. Techniques for colored noise are presented in [117, 131, 130].

2.4 Discussion

The benefit of the beamforming techniques is low computational complexity. However the estimates obtained are not generally consistent which is a serious drawback. The subspace methods have relatively low computational complexity and in the case of uncorrelated signals they have good performance. The limitation of the subspace methods is their performance reduction for correlated or even coherent signals. For ULAs, of course, the spatial smoothing preprocessing technique can be applied.

The ML methods can deal with coherent signals without any modifications. The SML method is more accurate than the DML method in Gaussian noise. The problem of the ML methods is the multidimensional search needed. WSF has the same asymptotic behavior as the SML but with a reduced computational complexity.

Note that the properties have been generally discussed under Gaussian assumptions. Gaussian noise is crucial for efficient methods such as SML or WSF. On the other hand the beamforming and subspace methods which are less efficient should be more robust to deviations from Gaussian noise.

In all the reviewed methods the sample covariance matrix is used for estimating DOAs. It is well known that the sample covariance matrix is extremely sensitive to deviations from Gaussian noise. If the data are non-Gaussian it may be a poor estimator of the true covariance matrix resulting in unreliable DOA estimates. The next chapter deals with robust estimation of the covariance matrix.

Chapter 3

Robust estimation of the covariance matrix

The mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ are natural parameters in the general M -variate Gaussian case. If the observations $\mathbf{x}_1, \dots, \mathbf{x}_N$ come from a $N_M(\boldsymbol{\mu}, \Sigma)$ -distribution (real or complex), the sample mean vector

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$

and the sample covariance matrix

$$S = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^H$$

are maximum likelihood estimators of $\boldsymbol{\mu}$ and Σ . Unfortunately, these estimators are extremely sensitive to deviations from the Gaussian assumption. This sensitivity is demonstrated in Fig. 3.1 where 50% tolerance ellipses obtained from these estimates are drawn for original and contaminated data. The ellipses enclose half of the data samples and their orientation and shape provide information about the covariance structure of the underlying distribution. The contaminated data is obtained by replacing one observation of the original data by an outlier. The outlying observation influences the sample covariance matrix significantly. As a consequence the shape and the orientation of the tolerance ellipse change.

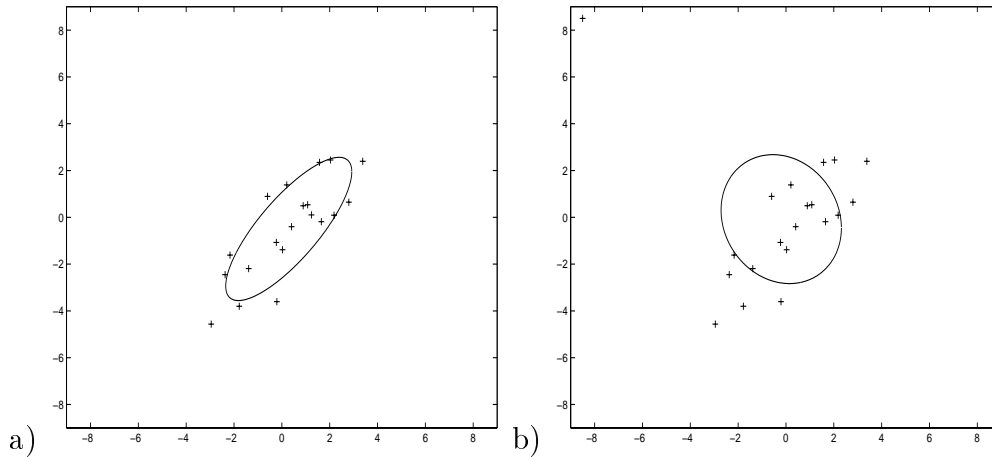


Figure 3.1: a) 50% tolerance ellipse from the original data. b) Tolerance ellipse after one observation is replaced by an outlier.

When data are non-Gaussian, robust estimators should be considered. In this chapter we focus on robust estimators of the covariance matrix. The chapter starts with an introduction to tools for measuring robustness and proceeds to a review of existing robust covariance matrix estimators. The remainder of the chapter deals with covariance matrix estimation techniques based on multivariate sign and rank concepts.

3.1 Tools for measuring robustness

An estimator is robust if it is insensitive to deviations from assumed conditions. The deviations may take the form of *outliers*, observations that do not follow the pattern of the majority of the data. Other sources of deviations are model class selection errors and incorrect assumptions on the measurement noise distribution. In this section we discuss two important tools for measuring robustness: the *breakdown point* and the *influence function*. Roughly speaking, the breakdown point gives the maximum fraction of bad outliers the estimator can cope with [42]. It therefore reflects *quantitative robustness* of an estimator. The influence function describes *qualitative robustness*. Its importance lies in a heuristic interpretation: it measures the effect of an infinitesimal contamination on the estimate.

3.1.1 Breakdown point

The concept of the breakdown point was introduced by Hodges [40] in the context of one-dimensional location estimation. A more general definition was given by Hampel [34].

The definition given by Donoho and Huber [28] will now be considered. Let $X_0 = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a set of M -variate data with T_N an estimator. The estimate given data X_0 is $T_N(X_0)$. Let $T_N(X_k)$ denote an estimate from data where k observations of X_0 are replaced by arbitrary values. The finite sample replacement breakdown point $\epsilon^*(T_N, X_0, N)$ is the smallest fraction $\frac{k}{N}$ that causes an estimator to *break down*. For location estimator, a formal definition can be given by

$$\epsilon^*(T_N, X_0, N) = \min_{1 \leq k \leq N} \left\{ \frac{k}{N} \mid \sup_{X_k} \{ \|T_N(X_0) - T_N(X_k)\| \} = \infty \right\} \quad (3.1)$$

where the supremum is taken over all possible X_k and $\|\cdot\|$ is the Euclidean distance. In this case the breakdown occurs when the Euclidean distance of the difference of the estimates (the bias caused by the replaced observations) tends to infinity. For the sample mean, for example, the breakdown point is equal to $\frac{1}{N}$.

For covariance matrix estimator, the concept of breakdown may be considered using its eigenvalue decomposition. Let $\hat{\Sigma}$ be an estimator of the covariance matrix Σ of full rank. Let $\mathbf{u}_1, \dots, \mathbf{u}_M$ denote the eigenvectors of Σ ordered by its eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$. Similarly let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_M$ be the eigenvalues of $\hat{\Sigma}$ and $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M$ the corresponding eigenvectors. The covariance matrix Σ may now be given as

$$\Sigma = U \Lambda U^H = \lambda U C U^H, \quad (3.2)$$

where $U = [\mathbf{u}_1, \dots, \mathbf{u}_M]$ is the unitary matrix of eigenvectors, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_M)$ is the diagonal matrix of the corresponding eigenvalues, C is a diagonal matrix of the normalized eigenvalues c_i ($\prod_{i=1}^M c_i = \det(C) = 1$) and $\lambda^M = \det(\Sigma)$ is the Wilks generalized variance. Note that $\Lambda = \lambda C$. As Bensmail and Celeux [5], we use the terms *scale*, *shape* and *orientation* for the items λ , C and U respectively.

As discussed in paper I, the decomposition (3.2) allows description of the breakdown of $\hat{\Sigma}$ in several different ways:

1. The scale measured by the Wilks generalized variance $\det(\hat{\Sigma})$ or by the trace $\text{Tr}\{\hat{\Sigma}\}$ may increase over all bounds.
2. The condition number (shape),

$$\text{Cond}(\hat{\Sigma}) = \frac{\hat{\lambda}_1}{\hat{\lambda}_M}$$

i.e, the ratio of the largest and smallest eigenvalue, may grow very large in the presence of outliers.

3. Influential observations may change the “ordered” eigenvectors and consequently drastically change the subspace spanned by the s first or s last columns of $\hat{U} = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M]$.
4. The coordinate system for independent coordinates (given by orientation \hat{U}) may change.

Lopuhaä and Rousseeuw [63] define the finite sample replacement breakdown point for covariance matrix estimators as

$$\epsilon^*(T_N, X_0, N) = \min_{1 \leq k \leq N} \left\{ \frac{k}{N} \mid \sup_{X_k} D(T_N(X_0), T_N(X_k)) = \infty \right\} \quad (3.3)$$

where the supremum is taken over all possible X_k ,

$$D(T(X_0), T(X_k)) = \max \left\{ |\lambda_1(T(X_0)) - \lambda_1(T(X_k))|, |\lambda_M(T(X_0))^{-1} - \lambda_M(T(X_k))^{-1}| \right\}, \quad (3.4)$$

and $\lambda_1(T(\cdot))$ and $\lambda_M(T(\cdot))$ denote the largest and smallest eigenvalue, respectively. This definition is related to the items 1 and 2 above. For simultaneous definition of breakdown of multivariate location and covariance matrix estimates, see [113]. It should be noted that all the existing definitions for the breakdown point of covariance matrix estimators use only the eigenvalues to determine breakdown. However, the direction of eigenvectors is crucial in many multivariate procedures such as subspace estimation and principal component analysis.

Another frequently used robustness measure is the (asymptotic) contamination breakdown point. It is related to a contamination model $F = (1 - \epsilon)F_0 + \epsilon G$ where F_0 is the

nominal distribution, $\epsilon \in [0, 1)$ and *contamination distribution* G ranges over all distributions. If T_N is a consistent estimator denote by $T(F)$ its limit when the underlying distribution is F . The contamination breakdown point is then defined as a smallest contamination probability ϵ that makes $T(F)$ to break down. In case of covariance matrices the breakdown is usually defined using the condition number. See e.g. [36, 67].

3.1.2 Influence function

The influence function [35, 36] is a standard tool for characterizing the qualitative robustness of an estimator. The influence function is essentially the first derivative of the functional version of an estimator. Let F be an M -variate distribution function and $T(F)$ a statistical functional corresponding to a consistent estimator T_N (the limit of T_N when the underlying distribution is F). The influence function of T at F is defined as

$$IF(\mathbf{x}, T; F) = \lim_{\epsilon \rightarrow 0} \frac{T((1 - \epsilon)F + \epsilon\Delta_{\mathbf{x}}) - T(F)}{\epsilon}$$

for those \mathbf{x} where the limit exists. $\Delta_{\mathbf{x}}$ is a probability measure which puts mass 1 at the point \mathbf{x} . A robust estimator should have an influence function that is bounded and continuous. When the influence function is bounded, an outlier cannot have an arbitrarily large influence on the estimate. Continuity guarantees that small changes in data cause only small changes in the parameter estimate. The empirical influence function is obtained by using an empirical distribution and contaminating it with an outlier whose location is varied in an M -dimensional space.

A covariance matrix estimator given by functional $T(\mathbf{x})$ ($\mathbf{x} \sim F$) is said to be *affine equivariant* if $T(B\mathbf{x} + \mathbf{b}) = BT(\mathbf{x})B^H$ (B is an $M \times M$ matrix of full rank and \mathbf{b} is an M -vector). Note that in case of location estimator T the affine equivariance is defined as $T(B\mathbf{x} + \mathbf{b}) = BT(\mathbf{x}) + \mathbf{b}$. For real valued affine equivariant covariance matrix estimators the influence function of eigenvalues or eigenvectors can be obtained from the influence function of the covariance matrix estimator [23].

It is often desirable to describe the effect of an outlier to a covariance matrix estimator with a scalar quantity instead of a matrix function. In paper *I*, various different tools for

this kind of comparisons are introduced. Let $\hat{\Sigma}$ be a covariance matrix estimate obtained from the uncontaminated data and let $\tilde{\Sigma}$ be an estimate obtained from the contaminated data. Denote the eigenvalues of $\hat{\Sigma}$ by $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_M$ and the corresponding eigenvectors by $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M$. Similarly let $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_M$ be the eigenvalues of $\tilde{\Sigma}$ and let $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_M$ be the corresponding eigenvectors. The eigenvector matrices of $\hat{\Sigma}$ and $\tilde{\Sigma}$ are $\hat{U} = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M]$ and $\tilde{U} = [\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_M]$, respectively.

As stated earlier, deviations caused by outliers to the estimated covariance matrix are conveniently described using eigenvalue decomposition. Changes in scale and shape may be described using product or the sum of the eigenvalues, for instance with $\det(\tilde{\Sigma})/\det(\hat{\Sigma})$ or traces $\text{Tr}(\tilde{\Sigma})/\text{Tr}(\hat{\Sigma})$. Perturbations in the shape may be captured using the whole spectrum of eigenvalues and the matching distance metric (see [101]):

$$md(\hat{\Sigma}, \tilde{\Sigma}) = \min_{\tau} \max_i (|\tilde{\lambda}_{\tau_i} - \hat{\lambda}_i|)$$

where $\tau = (\tau_1, \dots, \tau_M)$ is taken over all permutations of $(1, 2, \dots, M)$. If the perturbation is small the matching distance will be small and matching pairs of eigenvalues are clearly found.

Change in the condition number indicates change in the eigenvalue spread and how ill-conditioned the covariance matrix has become due to contamination. Also the change in the ratio of the geometric and arithmetic mean of the eigenvalues could be considered.

Perturbation of the orientation may be described in terms of the directions of the eigenvectors. One may investigate the perturbation effects on all the eigenvectors or on a subspace spanned by a subset of eigenvectors. Typically interesting eigenvectors are those corresponding to either the s largest or s smallest eigenvalues of $\hat{\Sigma}$. Let \hat{U}_s be a subset of eigenvectors from \hat{U} and \tilde{U}_s the corresponding matrix from \tilde{U} . Then $|\det(\tilde{U}_s^H \hat{U}_s)|$ may be used to quantify the change in the subspace spanned by s columns of \hat{U} . This quantity approaches unity when the subspaces come perfectly aligned. A more intuitive quantity is perhaps obtained by describing the change in the basis vectors of the subspace in terms of singular values of $\tilde{U}_s^H \hat{U}_s$. The canonical angles between the eigenvectors are obtained by $\cos^{-1}(\lambda_i)$ where λ_i are the singular values of $\tilde{U}_s^H \hat{U}_s$. Subspace \hat{U}_s and perturbed subspace \tilde{U}_s are close if the largest canonical angle is small.

In paper *I*, sensitivity plots for covariance matrix estimators were introduced based on the tools considered above. The concept of the sensitivity plot may be illustrated by using the concept of shape (the eigenvalue spread) as an example. Let $\hat{\Sigma}$ be the estimate of the covariance matrix calculated from the original data $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$. The change in the condition number of the covariance matrix estimate calculated from the perturbed data may be illustrated by plotting

$$SP(\mathbf{x}, X, \hat{\Sigma}) = \frac{\text{Cond}(\tilde{\Sigma})}{\text{Cond}(\hat{\Sigma})}$$

where $\tilde{\Sigma}$ is the covariance matrix estimate calculated from the perturbed data $X' = \{\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{x}\}$. Similarly, the changes in orientation and scale may be plotted using either the difference between the true and perturbed values or their ratios, depending on which is more appropriate for the quantity of interest.

Examples of this type of sensitivity plots (averages over 50 samples) for the sample covariance matrix are provided in Fig. 3.2. The plots were created in the following way. Independent samples of sizes 100 were drawn from the real-valued bivariate Gaussian distribution with symmetry center $\boldsymbol{\mu} = (0, 0)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}.$$

Each of the samples was perturbed with the same multivariate outlier. The changes between the original and perturbed estimate were quantified using two criteria, the proportional change in the condition numbers (shape) and the change in the direction of the first eigenvector (orientation). As seen from the figure, the influence of one additional observation on the sample covariance matrix is unbounded.

3.2 A review of robust estimators of the covariance matrix

In this section a review of robust covariance matrix estimation techniques is given (for another recent review, see [67]). The methods considered are M -estimators, S -estimators,

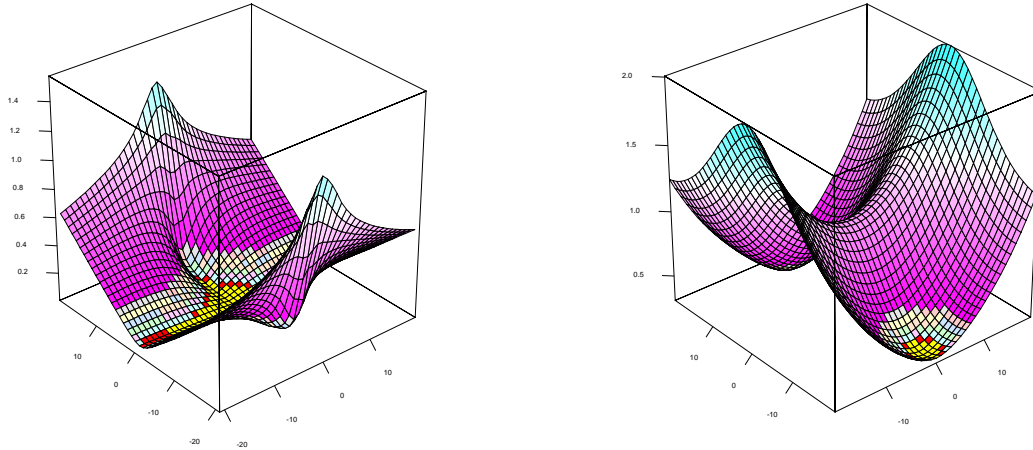


Figure 3.2: Sensitivity plots for the sample covariance matrix. Used criteria are the difference in the direction of the first eigenvector (left) and the ratio of the condition numbers (right). The influence of one additional observation on the sample covariance matrix is unbounded.

Minimum Volume Ellipsoid (MVE), Minimum Covariance Determinant (MCD) and estimates based on projections. Properties of the estimators discussed are the breakdown point, influence function and efficiency. In this section it is assumed that the M -dimensional observations are in *general position*, which means that no more than M points of the data lie in any $(M - 1)$ -dimensional subspace. This assumption is made for theoretical reasons.

The methods introduced are for real-valued data. For complex valued M -dimensional observations, the covariance matrix can be estimated by combining the real and imaginary parts of the observations into $2M$ -dimensional real observation vectors, and then estimating the covariance matrix for these pseudo observations. A final estimate can be formed from the elements of the real matrix as described in paper V.

3.2.1 M -estimation

M -estimation is a generalization of Maximum Likelihood estimation principle. Here the M -estimates for multivariate location and scatter are considered. The reason why the term scatter is used instead of covariance matrix will be evident later. We start from an M -variate random variable \mathbf{y} which has a *spherically symmetric distribution* [45, 30, 29] for which a density function exists. The density function of \mathbf{y} is of the form $g(\mathbf{y}^T \mathbf{y})$ for some nonnegative function $g(\cdot)$ of scalar variable, i.e. it depends only on the Euclidean distance of \mathbf{y} from the origin. Let B be an $M \times M$ matrix of full rank and $\boldsymbol{\mu}$ an M -vector. The density function of the linearly transformed variable $\mathbf{x} = B\mathbf{y} + \boldsymbol{\mu}$ can be given as

$$f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \det(\Sigma)^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})), \quad (3.5)$$

where $\Sigma = BB^T$. Distributions of this form belong to the family of *elliptically symmetric* distributions. The matrix Σ is the ordinary covariance matrix only if \mathbf{y} has a covariance matrix and it is equal to I . Therefore Σ is called a scatter matrix or pseudo covariance matrix.

M -estimators are closely related to the elliptically symmetric density given in (3.5). Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be a data set in \mathbb{R}^M and let $\rho(s)$ be a given function of s . Denote the set of positive definite symmetric $M \times M$ matrices by $PDS(M)$. The goal is to find $\hat{\boldsymbol{\mu}} \in \mathbb{R}^M$ and $\hat{\Sigma} \in PDS(M)$ minimizing the objective function

$$L(\boldsymbol{\mu}, \Sigma) = \sum_{i=1}^N \rho[(\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu})] + \frac{1}{2} \log[\det(\Sigma)]. \quad (3.6)$$

When $\exp\{\rho(\mathbf{x}^T \mathbf{x})\}$ is integrable over \mathbb{R}^M , (3.6) can be considered as a negative log likelihood function for elliptically symmetric distribution. If ρ is differentiable, then setting derivative of (3.6) with respect to $\boldsymbol{\mu}$ and Σ to 0 yields the estimation equations

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^N w(s_i) \mathbf{x}_i}{\sum_{i=1}^N w(s_i)} \quad (3.7)$$

$$\hat{\Sigma} = \sum_{i=1}^N w(s_i) (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T (\mathbf{x}_i - \hat{\boldsymbol{\mu}}) \quad (3.8)$$

where $w(s) = 2\rho'(s)$ and $s_i = (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})$.

M -estimates for $\boldsymbol{\mu}$ and Σ are defined as a generalization of (3.7) and (3.8). The estimates are defined to satisfy

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^N w_1(s_i) \mathbf{x}_i}{\sum_{i=1}^N w_1(s_i)} \quad (3.9)$$

$$\hat{\Sigma} = \frac{\sum_{i=1}^N w_2(s_i) (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T}{\sum_{i=1}^N w_3(s_i)}. \quad (3.10)$$

where w_1 , w_2 and w_3 are functions from $[0, \infty)$ to $[0, \infty)$.

M -estimators are affine equivariant. For sample versions this means that if $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are M -estimates for original observations, the M -estimates for affine transformed observations $B\mathbf{x}_i + \mathbf{b}$ (B is an $M \times M$ matrix of full rank, \mathbf{b} an M -vector) are given by $B\hat{\boldsymbol{\mu}} + \mathbf{b}$ and $B\hat{\Sigma}B^T$. The existence and uniqueness of the solutions to (3.9)-(3.10) depends on the weighting functions and the data available. Assuming the location parameter $\boldsymbol{\mu}$ known, Huber [43] proved the existence and uniqueness of solutions for suitable weighting functions. Kent and Tyler [49] discussed the same issues for M -estimates defined directly to maximize (3.6) and included discussion of joint estimation of $\boldsymbol{\mu}$ and Σ . For existence results see also [112, 36]. When unique solutions exist, the resulting estimates are consistent and asymptotically normal [43, 68]. The theoretical conditions for the proofs are somewhat complicated and will not be discussed here.

Equations (3.9)-(3.10) can be used as the basis of iterative algorithms for numerical computation of M -estimates. More sophisticated ideas are discussed in [68, 43]. Convergence of iterative algorithms was considered in [49]. An example of weight functions can be given by $w_3(s) = 1$ and $w_i(s) = \psi_i(s)/s$, for $i = 1, 2$, where $\psi_1(s) = \psi_H(\sqrt{s}, k)$ and $\psi_2(s) = \psi_H(s, k^2)$. The function $\psi_H(s, k) = \min\{s, \max\{s, -k\}\}$ is known as Huber's psi-function. Discussion of various weighting functions can be found, for example, in [43, 13].

The influence function of the M -estimators is bounded for suitably chosen weight functions [43]. Unfortunately the asymptotic contamination breakdown point of M -estimators is shown to be at most $1/(M + 1)$ [68] (for the definition of breakdown given in the reference). This means that M -estimates become more sensitive to the outlying observations as the dimension of the data grows. To overcome this problem, Kent and Tyler [50] introduced

an additional constraint to the estimation equations (3.7)-(3.8). The resulting estimates are called constrained M -estimates and can attain a highest contamination breakdown point of $1/2$.

3.2.2 MVE and MCD

Rousseeuw [89] introduced two extremely robust affine equivariant methods for multivariate location and covariance matrix estimation: the Minimum Volume Ellipsoid (MVE) and the Minimum Covariance Determinant (MCD). These estimators are based on the idea of using only a fraction of the available data in estimation task.

The MVE estimate for location is the center of the ellipsoid covering at least h points of data. The scatter estimate is the shape matrix determining this ellipsoid. Mathematical derivation can be given as follows [63]. Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be M -variate data. The MVE estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma} \in PDS(M)$ for location and scatter minimize the determinant of Σ subject to

$$\#\{i | (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \leq c^2\} = h.$$

The number c is a fixed constant and has no influence on the value of $\hat{\boldsymbol{\mu}}$. However, the choice of c determines the magnitude of $\hat{\Sigma}$. The value of c can be chosen with agreement of the underlying distribution in order to obtain a consistent covariance estimate. For example, if $h = \lfloor (n + p + 1)/2 \rfloor$ and the data are assumed to come from Gaussian distribution, c^2 should be the median of a χ^2 distribution with M degrees of freedom.

The breakdown point of MVE estimates depends naturally on the choice of h . If $h = \lfloor (n + p + 1)/2 \rfloor$, the finite sample replacement breakdown point (as defined in equations (3.1)-(3.3)) will attain the maximum value for affine equivariant estimators of $\lfloor (N - M + 1)/2 \rfloor / 2$ [63]. The influence function of MVE estimates has not been considered in statistical literature. The main weakness of the MVE estimator is its poor efficiency when h is close to $1/2$ [25]. Therefore the MVE estimates are usually used to obtain robust initial estimates. The final estimates for location and covariance are then obtained applying one step reweighting. The weighting is done by using robust Mahalanobis-type distances obtained from the MVE estimates. It has been shown that with an appropriate weighting function the breakdown

point of the final estimate is the same as the breakdown point of the initial estimates [63].

Several algorithms exist for computation of MVE estimates [93, 38, 122, 19, 46, 1, 21]. The most commonly used are probabilistic algorithms based on drawing a large number of small subsamples from the data.

For M -variate data $\mathbf{x}_1, \dots, \mathbf{x}_N$ the MCD estimate for location is the arithmetic mean of h points for which the determinant of the sample covariance matrix is minimal [89, 92]. The covariance matrix estimate is the sample covariance matrix of these points. MCD estimators are more efficient than MVE estimators and have the same high robustness [11, 22]. Therefore MCD estimates are preferred over MVE estimates. The influence function of an MCD estimator is naturally bounded [22]. Recently a fast algorithm for numerical computation of MCD estimates was developed [91]. For other algorithms see [51, 39, 122]. As in the case of the MVE estimators the MCD estimates are usually used to obtain robust Mahalanobis-type distances. These distances are then used to form a weighted mean and a weighted sample covariance matrix of the data.

3.2.3 S -estimates

For data $\mathbf{x}_1, \dots, \mathbf{x}_N$ in \mathbb{R}^M the S estimates for multivariate location and scatter are defined as a pair $(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) \in \mathbb{R}^M \times PDS(M)$ minimizing $\det(\hat{\Sigma})$ subject to

$$\frac{1}{N} \sum_{i=1}^N \rho \left[\{(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \hat{\Sigma}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})\}^{1/2} \right] = b.$$

The MVE estimator can be obtained as a special case of an S -estimator by letting $Nb = N - h$ and $\rho(\cdot) = 1 - 1_{\{-c, c\}}$, where $1_{\{-c, c\}}$ is the indicator function over $[-c, c]$. To ensure asymptotic normality and a high breakdown point the following conditions are usually assumed of ρ :

- ρ is symmetric, twice continuously differentiable, and $\rho(0) = 0$.
- there exists a constant $c > 0$ such that ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$.

A typical example of such a ρ function is given by

$$\rho(y) = \begin{cases} \frac{y^2}{2} - \frac{y^4}{2c^2} + \frac{y^6}{6c^4}, & \text{for } |y| \leq c, \\ \frac{c_0^2}{6}, & \text{for } |y| > c. \end{cases}$$

The constant $0 < b < \sup\{\rho\}$ can be chosen in agreement of the underlying probability distribution. If the data $\mathbf{x}_1, \dots, \mathbf{x}_N$ is assumed to be a random sample from elliptically symmetric distribution with density (3.5), it is natural to choose $b = E\{\rho((\mathbf{x} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))\}$, where \mathbf{x} is distributed according to assumed density. The choice of the constant c effects both the breakdown point and asymptotic variance of the S -estimates. By choosing c depending on the number of observations and supremum of ρ it is possible to obtain the maximal finite sample replacement breakdown point of affine equivariant estimators [62, 63]. However, it is not possible to combine small asymptotic variance and high breakdown point [62]. The influence function of S -estimators with ρ function satisfying the conditions above is bounded [62].

The S -estimators are closely related to the M -estimates. In fact the S -estimates satisfy the M -estimation equations (3.9)-(3.10) for weights $w_1(s) = \rho'(s^{1/2})/s^{1/2}$, $w_2(s) = w_1(s)$ and $w_3(x) = M^{-1}s^{1/2}\rho'(s^{1/2}) - \rho(s^{1/2}) + b_0$ [62]. The S -estimates, however, have a breakdown point which is independent of the dimension of the data. This difference is due to the fact that S -estimates found by the minimization problem, which is not equivalent to solving estimation equations (3.9)-(3.10) with these weights. Algorithms for numerical computation of the S -estimators appear in [95, 14].

3.2.4 Estimates based on projections

The Stahel-Donoho estimate

The estimator defined independently by Stahel [99] and Donoho [27] was the first robust affine equivariant estimator for multivariate location and scatter having a high breakdown point for any dimension. Estimators use an idea that an multivariate outlier should also be an outlier at least in one univariate projection of the data. Robust univariate scale and location statistics are used to measure “outlyingness” of an observation in all possible

directions. The outlyingness measure is then used to form a weighted mean and a weighted sample covariance matrix for the data.

Let $\mu(\cdot)$ and $\sigma(\cdot)$ be univariate affine equivariant location and scale statistics¹. Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a data set of N observations in \mathbb{R}^M . Define for any $\mathbf{y} \in \mathbb{R}^M$ the “outlyingness” r :

$$r(\mathbf{y}, X) = \sup_{\mathbf{a}} \left\{ \frac{|\mathbf{a}^T \mathbf{y} - \mu(\mathbf{a}^T X)|}{\sigma(\mathbf{a}^T X)} \right\}$$

where the supremum is over $\mathbf{a} : \|\mathbf{a}\| = 1$. Usually $\mu(\cdot)$ is taken to be the median and $\sigma(\cdot)$ is taken to be the Median Absolute Deviation (MAD). For a univariate data x_1, \dots, x_N , the MAD is defined as

$$\text{MAD}(X) = c \cdot \text{med}_{i=1, \dots, N} \{|x_i - \text{med}\{x_1, \dots, x_N\}|\}$$

where c is a consistency correction constant (for the Gaussian distribution $c \approx 1.4826$ makes the MAD consistent towards the standard deviation). See, for example, [43]. Let $w(\cdot)$ be a positive weighting function. The Stahel-Donoho estimator of location and scatter $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ is defined as

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^N w_i \mathbf{x}_i}{\sum_{i=1}^N w_i}$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{\sum_{i=1}^N w_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T}{\sum_{i=1}^N w_i},$$

with $w_i = w(r(\mathbf{x}_i, \mathbf{X}))$. If w is continuous and $w(r)$ and $r^2 w(r)$ are bounded for $r \geq 0$ and if μ and σ have asymptotic breakdown points of $1/2$, the asymptotic contamination breakdown point of $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ is $1/2$ in continuous multivariate models [36]. The finite sample breakdown point of these estimators was considered in [113]. The estimates are consistent (of order \sqrt{N}) under some general conditions [66]. The numerical computation of the Stahel-Donoho estimates is complex. An approximative algorithm based on subsampling was proposed by Stahel [99].

¹For univariate data $X = \{x_1, \dots, x_N\}$ and $a, b \in \mathbb{R}$ $\mu(aX + b) = a\mu(X) + b$, $\sigma(aX + b) = |a|\sigma(X)$.

***P*-estimates**

Another affine equivariant estimator based on projections was proposed by Maronna et al. [65]. Let B be an $M \times M$ matrix satisfying $B^T B = \Sigma^{-1} \in PDS(M)$ and \mathbf{x} a random variable with covariance matrix Σ . The estimator is based on the idea that for any $\mathbf{a} \in R^M$ with $\|\mathbf{a}\| = 1$, one has $\text{var}(\mathbf{a}^T B \mathbf{x}) = 1$. That is, B induces a transformation of the data such that the variance is the same in all directions. The proposed method is based on replacing the variance by a robust univariate dispersion estimate $\sigma(\cdot)$ such as the MAD. A *P-estimate* of scatter is defined as $\hat{\Sigma} = (B^T B)^{-1}$ where B minimizes

$$\frac{\sup_{\mathbf{a}} \sigma(\mathbf{a}^T B \mathbf{x})}{\inf_{\mathbf{a}} \sigma(\mathbf{a}^T B \mathbf{x})}$$

with $\mathbf{a} \neq \mathbf{0}$. The numerical computation of *P*-estimates is especially difficult and therefore the estimator has not drawn much attention.

For still another projection pursuit based approach see [43, 61]. These methods first compute a robust estimate of the first eigenvector by finding the direction which yields the maximum univariate robust estimate of variance. The direction of the second eigenvector is orthogonal to the first one and again yields the maximum robust univariate variance estimate. The remaining eigenvalues and eigenvectors are found sequentially in similar fashion. This estimator is called the PPS estimator. The PPS estimator can obtain an asymptotic breakdown point of $1/2$ if the robust estimate of the variance has an asymptotic breakdown point of $1/2$ [61].

Let $\hat{\Sigma}$ denote the PPS estimate computed from data $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. The PPS estimate is *orthogonal equivariant*. That is, for any $M \times M$ orthogonal matrix U , the PPS estimate from data UX is $U\hat{\Sigma}U^T$. In general, PPS estimates are not affine equivariant. However, if the underlying distribution is an elliptic one, then in the asymptotic sense they are [61]. An approximative algorithm for PPS estimates is given in [24].

3.3 Robust estimation using nonparametric statistics

In this section, techniques for estimating the covariance matrix using spatial signs and ranks are considered. First, recall the univariate sign and rank concepts. Let x_1, \dots, x_N be an

univariate real-valued data set. The univariate sign function $S(x) = \text{sign}(x)$ is 1, 0, -1 as $x > 0, = 0, < 0$ and the centered rank function is $R(x) = \frac{1}{N} \sum_{i=1}^N S(x - x_i)$. Note that $R(x)$ is the derivative function of the criterion function for the median $D(x) = \frac{1}{N} \sum_{i=1}^N |x - x_i|$.

For a real M -vector \mathbf{x} , the *spatial sign function* [16, 71] is defined as

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0} \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

where $\|\mathbf{x}\| = (\mathbf{x}^T \mathbf{x})^{1/2}$. The spatial sign of \mathbf{x} is a unit length vector to the direction of \mathbf{x} and hence a natural generalization of the univariate sign function.

Let now $\mathbf{x}_1, \dots, \mathbf{x}_N$ be a real valued M -variate data set. The spatial rank function is defined as

$$\mathbf{R}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \mathbf{S}(\mathbf{x} - \mathbf{x}_i).$$

Note that $\mathbf{R}(\mathbf{x})$ is the gradient of the criterion function $D(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{x} - \mathbf{x}_i\|$. The spatial median $\mathbf{M}(X)$ minimizes $D(\mathbf{x})$ or is the solution of equation $\mathbf{R}(\mathbf{x}) = \mathbf{0}$. See [16, 17, 71]. For complex valued data the spatial sign and rank functions and spatial median are defined by using $\|\mathbf{x}\| = (\mathbf{x}^H \mathbf{x})^{1/2}$ in above. In what follows we discuss the covariance properties of the spatial sign and rank in the general complex case.

The sample Sign Covariance Matrix (SCM) denoted by S_1 , the sample Tau Covariance Matrix (TCM) denoted by S_2 and the sample Rank Covariance Matrix (RCM) denoted by S_3 are defined as

$$S_1 = \frac{1}{N} \sum_{i=1}^N \mathbf{S}(\mathbf{x}_i) \mathbf{S}^H(\mathbf{x}_i), \quad S_2 = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \mathbf{S}(\mathbf{x}_i - \mathbf{x}_j) \mathbf{S}^H(\mathbf{x}_i - \mathbf{x}_j)$$

and

$$S_3 = \frac{1}{N} \sum_{i=1}^N \mathbf{R}(\mathbf{x}_i) \mathbf{R}^H(\mathbf{x}_i),$$

respectively.

To define the corresponding theoretical concepts, let $\mathbf{x}_1, \mathbf{x}_2$ and \mathbf{x}_3 be i.i.d. M -variate random variables with the distribution F . Then the SCM, TCM and RCM for the distribution F are

$$\Sigma_1 = E_F\{\mathbf{S}(\mathbf{x}_1) \mathbf{S}^H(\mathbf{x}_1)\}, \quad \Sigma_2 = E_F\{\mathbf{S}(\mathbf{x}_1 - \mathbf{x}_2) \mathbf{S}^H(\mathbf{x}_1 - \mathbf{x}_2)\}$$

and

$$\Sigma_3 = E_F\{\mathbf{S}(\mathbf{x}_1 - \mathbf{x}_2)\mathbf{S}^H(\mathbf{x}_1 - \mathbf{x}_3)\},$$

respectively. For i.i.d. observations the sample versions converge w.p.1 (with probability 1) to the theoretical matrices. This result is proven for the SCM and the TCM in paper V. The technique used in these proofs can also be applied to obtain the result for the RCM.

Because the Euclidean distance is invariant under unitary transformations, the spatial sign function is *rotation equivariant*, i.e. $\mathbf{S}(U\mathbf{x}) = U\mathbf{S}(\mathbf{x})$ for any unitary matrix U . Therefore, if the SCM, TCM and RCM for the distribution of \mathbf{x} are Σ_1 , Σ_2 and Σ_3 , then the SCM, TCM and RCM for the distribution of $U\mathbf{x}$ are $U\Sigma_1U^H$, $U\Sigma_2U^H$ and $U\Sigma_3U^H$. These properties naturally hold also for the sample versions. The spatial sign function, however, is not scale equivariant, i.e. it is not true that $\mathbf{S}(D\mathbf{x}) = D\mathbf{S}(\mathbf{x})$ for all diagonal matrices D .

There is a close connection between the eigenvectors of the covariance matrix and the eigenvectors of the SCM, TCM and RCM for a large class of multivariate distributions. In order to show this relation, we start by giving definitions for reflection and permutation invariance of multivariate distributions. We say that an $M \times M$ matrix G is a *reflection matrix* if it is a diagonal matrix with diagonal elements ± 1 . Moreover, an $M \times M$ matrix Q is a *permutation matrix* if it is obtained by permuting the rows (or columns) of the $M \times M$ identity matrix. A distribution of \mathbf{z} is said to be *reflection invariant* if $G\mathbf{z} \sim \mathbf{z}$ ($G\mathbf{z}$ and \mathbf{z} have identical distributions) for all reflection matrices G and is *permutation invariant* if $Q\mathbf{z} \sim \mathbf{z}$ for all permutation matrices Q .

Theorem 1 *Let \mathbf{z} be an M -variate random variable with a reflection and permutation invariant distribution. Consider a random variable*

$$\mathbf{x} = UD\mathbf{z}$$

where $D = \text{diag}\{d_1, \dots, d_M\}$, $|d_1| \geq \dots \geq |d_M| > 0$ and U is a unitary matrix. Let the covariance matrix and SCM of \mathbf{x} be Σ and Σ_1 , respectively. Then

$$\Sigma = U\Lambda U^H \quad \text{and} \quad \Sigma_1 = U\Lambda_1 U^H$$

where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_M\}$ $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > 0$, and $\Lambda_1 = \text{diag}\{\lambda_{1,1}, \dots, \lambda_{1,M}\}$, $\lambda_{1,1} \geq \lambda_{1,2} \geq \dots \geq \lambda_{1,M} > 0$. Moreover $\lambda_i = \lambda_{i+1}$ if and only if $\lambda_{1,i} = \lambda_{1,i+1}$, or in other words the eigenvectors of Σ and Σ_1 ordered by their respective eigenvalues can be chosen to be the same.

Proof. See appendix A.

Let \mathbf{z}_1 and \mathbf{z}_2 be i.i.d. random variables with a reflection and permutation invariant distribution. Then also the random variable $\mathbf{z}_1 - \mathbf{z}_2$ has a reflection and permutation invariant distribution and we get the following corollary.

Corollary 1 *Let \mathbf{z} be a M -variate random variable with a reflection and permutation invariant distribution. Consider a random variable*

$$\mathbf{x} = U D \mathbf{z} + \mathbf{b}$$

where U and D are as in Theorem 1 and \mathbf{b} is a complex M -vector. Let the covariance matrix and the TCM of \mathbf{x} be Σ and Σ_2 , respectively. Then

$$\Sigma = U \Lambda U^H \quad \text{and} \quad \Sigma_2 = U \Lambda_2 U^H$$

where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_M\}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > 0$, and $\Lambda_2 = \text{diag}\{\lambda_{2,1}, \dots, \lambda_{2,M}\}$, $\lambda_{2,1} \geq \lambda_{2,2} \geq \dots \geq \lambda_{2,M} > 0$. Moreover $\lambda_i = \lambda_{i+1}$ if and only if $\lambda_{1,i} = \lambda_{1,i+1}$.

Let now \mathbf{z} , U , D and \mathbf{b} be as in Corollary 1. It is straightforward to show that the RCM of the distribution $U D \mathbf{z} + \mathbf{b}$ is $U \Lambda_3 U^H$, where Λ_3 is a diagonal matrix of eigenvalues (for real valued random variables, see [64]). Unfortunately, showing that the order of the eigenvectors is preserved is more complicated.

The class of distributions introduced in Corollary 1 is large and includes the circular complex Gaussian distribution and all complex elliptically symmetric distributions considered in [56] (with existing second order moments). It is, however, larger than the family considered in [56]. Real valued distributions belonging to the family introduced in Corollary 1 include all elliptically symmetric distributions with existing second order moments. In Theorem 1 and Corollary 1 the covariance matrix of \mathbf{x} was assumed to exist just to show

the relation between the eigenvectors. Because the spatial sign function is bounded, the SCM TCM and RCM exist for all multivariate distributions. Therefore, the results stated for the SCM, TCM and RCM are also valid for the constructions where the second order moments of \mathbf{z} are not defined, such as the multivariate Cauchy distribution.

The above results suggest that the sample TCM or RCM may be used to find estimates for the eigenvectors of the covariance matrix. An estimate of the covariance matrix can be constructed by combining these estimates with a robust estimate of the variance. The estimation strategy may be as follows:

1. Calculate the sample TCM or the sample RCM from the data $\mathbf{x}_1, \dots, \mathbf{x}_N$. Find the corresponding eigenvector estimates, that is, a matrix \hat{U} .
2. Estimate the marginal variances (eigenvalues, principal values) of

$$\hat{U}^H \mathbf{x}_1, \hat{U}^H \mathbf{x}_2, \dots, \hat{U}^H \mathbf{x}_N$$

using any scale equivariant univariate robust scale estimate (MAD, etc.). Write $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_M)$ for the estimates ($\hat{\lambda}_1$ corresponds to the first marginal etc.).

3. The covariance matrix estimate is

$$\hat{\Sigma} = \hat{U} \hat{\Lambda} \hat{U}^H.$$

If the sample SCM is employed in covariance matrix estimation, the data has to be first centered with respect to the spatial median of the data (see paper *I*). The estimates constructed using the sample SCM, TCM or RCM are rotation (orthogonal) equivariant, but not affine equivariant. Simulation results indicate that the eigenvector estimates based on the sample SCM, TCM or RCM have good efficiency and robustness properties. In paper *I*, simulation results concerning the sample TCM are presented. For SCM and RCM based methods, see [64]. Note that results regarding the breakdown properties for projection based methods introduced in [113, 61] are also valid for the estimators obtained using the steps above.

In paper *II* the use of TCM and RCM is demonstrated in different multichannel signal processing tasks. Robust covariance matrix estimates obtained from the sample RCM or

the sample TCM are used in RGB color image filtering, principal component analysis, discrete Karhunen-Loève transform and blind source separation problem.

3.4 Discussion

Probably the most applied robust estimates of the covariance matrix are the MVE and MCD estimates. One of the main reasons for their popularity is that algorithms for their numerical calculation are easily available. For example the S-PLUS software [98] has functions for both MVE and MCD based covariance matrix estimates. Examples where these estimates are used in various applications can be found from [90]. The major drawback of the MVE estimate is poor efficiency when compared to other methods. The MCD estimator has better efficiency and therefore it is preferred over MVE estimator.

The properties of the M , S , and reweighted MCD estimators are compared in [23]. The M -estimator and S -estimator are based on the weight functions given as an example when introducing the estimation methods in section 3.2. The results in [23] show that the efficiency of the reweighted MCD estimator is lower than the efficiency of M - or S -estimators. The authors recommend use of the S -estimator because it seems to be both efficient and robust. Their recommendation is motivated by both theoretical and simulation results.

The estimates based on projections have not received much attention in the research literature. A simulation study comparing the Stahel-Donoho estimates to MVE, S , and M -estimates is presented in [66]. The results of this study are extremely difficult to interpret, but the authors' conclusion is that the Stahel-Donoho estimator is preferable to the others. The weakness of the Stahel-Donoho estimates is in numerical computation.

Numerical computation of estimates based on spatial sign and rank concepts is straightforward. This is a benefit of these estimators when compared to methods already discussed in this section. When a whole covariance matrix estimate is formed, the lack of affine equivariance is a disadvantage. However, the eigenvector estimates obtained are rotation equivariant, which is the only natural requirement for such estimates. Note that the dis-

tribution family introduced in Theorem 1 and Corollary 1 does not generally cover the signal model (2.1) used in array signal processing. In the next chapter we show that the sample SCM and the sample TCM can, however, be used to estimate the signal and noise subspaces for a large family of symmetric noise distributions.

Chapter 4

Robust DOA estimation

Measurements of real-world channels reveal that noise appearing in many indoor and outdoor mobile communication channels is non-Gaussian. See [69, 60, 110, 126, 115, 54] and references therein. The array processing techniques introduced in chapter 2 rely heavily on the sample covariance matrix. Therefore these techniques often perform poorly when the Gaussian assumption is not valid.

In this chapter we consider robust DOA estimation and robust estimation of the number of signals. The techniques addressed are robust against heavy-tailed non-Gaussian noise. The chapter starts with a review of common noise models and robust DOA estimation algorithms. The remainder of the chapter states the main result of this dissertation. We show that the sample SCM and the sample TCM can be used to obtain convergent estimates for the signal and noise subspace basis vectors. These estimates can then be applied to construct robust estimates of DOAs and the number of signals. We also discuss how nonparametric statistics may be applied to frequency estimation, and show the theoretical motivation for the resulting techniques. The chapter ends with discussion.

4.1 Noise distribution families

Experimental measurements show that man-made interference has an impulsive nature that can not be modeled well by a Gaussian distribution [115]. Thus actual channel noise can deviate greatly from the typical Gaussian assumption. For example, measurements of

outdoor urban radio channels indicate that automobile ignition noise levels exceed those of typical thermal (Gaussian) noise, see [54] and references therein. Impulsive noise may be modeled with heavy-tailed distributions. In what follows we introduce noise distribution families that are used in robust statistical signal processing.

Probably the most common distribution family applied in robust estimation is ϵ -contamination family. This family contains distributions given by

$$\mathbf{y} = (1 - b)\mathbf{x} + b\mathbf{h},$$

where $b \sim \text{Bin}(1, \epsilon)$, \mathbf{x} is distributed according to a nominal noise distribution and \mathbf{h} is distributed according to an arbitrary contaminating distribution. When \mathbf{y} is used to model noise, \mathbf{x} and \mathbf{h} are usually assumed to zero mean Gaussian random variables with covariance matrices $\sigma_1^2 I$ and $\sigma_2^2 I$, $\sigma_1^2 \ll \sigma_2^2$. This is a special case of Gaussian mixture noise-model. Univariate complex-valued random variable belonging to this family has density

$$f(y) = \sum_{l=1}^L \frac{\epsilon_l}{\pi \sigma_l^2} \exp\left\{-\frac{yy^*}{\sigma_l^2}\right\} \quad (4.1)$$

with $\sum_{l=1}^L \epsilon_l = 1$.

The family of α -stable distributions [97] is also widely used. The characteristic function of a univariate complex isotropic symmetric α -stable ($S\alpha S$) distribution is

$$\rho(\omega) = \exp(-\gamma|\omega|^\alpha).$$

The smaller the characteristic exponent $\alpha \in [0, 2]$, the heavier the tails of the density. The case of $\alpha = 2$ corresponds to the univariate complex circular Gaussian distribution and the case of $\alpha = 1$ corresponds to the univariate complex Cauchy distribution. The positive valued scalar γ is the dispersion of the distribution. The dispersion plays a role analogous to that of the variance for second order processes.

Finally we consider the family of complex spherically symmetric distributions. A complex-valued M -variate random variable \mathbf{y} is said to have a complex spherically symmetric distribution if the distribution of the $2M$ real vector

$$\tilde{\mathbf{y}} = \begin{pmatrix} \text{Re}\{\mathbf{y}\} \\ \text{Im}\{\mathbf{y}\} \end{pmatrix}$$

is spherically symmetric [29], i.e.

$$G\tilde{\mathbf{y}} \sim \tilde{\mathbf{y}}$$

for any $2M \times 2M$ orthogonal matrix G . This implies that $H\mathbf{y} \sim \mathbf{y}$ for any $M \times M$ unitary matrix H . Consequently the covariance matrix of \mathbf{y} is a constant times identity matrix, if the second-order moments of the distribution exist.

4.2 Review of robust DOA estimation methods

One of the first papers considering robust DOA estimation was [60]. In the approach proposed by the authors the array output vector is modeled as a multivariate AR process. The parameters of the AR process are then estimated using univariate M -estimation and an estimate for the array output covariance matrix is formed from these parameters. The method also requires estimation of model order. In [59] another method based on univariate M -estimation was proposed. The authors considered real-valued data model and assumed the signal vector to be known and constant. The DOAs were estimated using univariate M -estimation.

Williams and Johnson [121] assumed that the signals are circular Gaussian and derived an M -estimator of the covariance matrix based on the least informative noise distribution in the ϵ -contamination model. They also discussed ML estimation of the covariance matrix when the array output covariance matrix has a Toeplitz structure.

DOA estimation based on α -stable processes [97] was considered in [110, 111]. In [110], the authors derived pseudo ML-estimates for DOAs when the signals are modeled as deterministic and the marginal components of the noise are assumed to be i.i.d. from univariate complex-valued Cauchy distribution. In [111] the marginal components of signal and noise vector were modelled as complex isotropic $S\alpha S$ random variables. The authors proposed an algorithm wherein the noise subspace is estimated using a covariation matrix and the DOAs are estimated using the MUSIC algorithm applied on obtained noise subspace estimate. The covariation matrix is identical to the covariance matrix when the signal and noise are modeled as Gaussian, i.e. $\alpha = 2$. When $\alpha < 2$, the covariation matrix is based on fractional

lower order moments. The resulting algorithm is called RObust Covariation-based MUSIC (ROC-MUSIC).

Yardimci et al. [126] used a generalization of least-squares estimation in order to achieve robustness. In their approach the signals are modeled as deterministic and DOA estimates and estimates for the signal amplitudes are obtained by minimizing the criterion

$$\sum_{n=1}^N \psi[||\mathbf{x}(n) - A(\boldsymbol{\theta})\mathbf{s}(n)||^2]$$

where $\psi(\cdot)$ is a weighting function defined on the positive real axis. Note that DML estimates (or LS-estimates) are obtained by setting $\phi(x) = x$. The authors discussed optimal choice of weighting functions for specific noise models including complex spherically symmetric distributed noise and the ϵ -contamination model.

Kozick and Sadler [54] considered ML estimation when the noise is modeled as Gaussian mixture noise. They assumed that the marginal components of the noise are i.i.d. according to the distribution (4.1) and used the SAGE algorithm [32]. Signals were modeled as deterministic. They also derived the Cramér-Rao bound for noise with i.i.d. complex spherically symmetric marginal components, and discussed estimation of the number of sources. In their simulation study they compared their algorithms to MUSIC and ROC-MUSIC algorithms and concluded that ROC-MUSIC algorithm has poor performance unless the sample size N is very large ($N \geq 2000$).

4.3 Robust estimation using nonparametric statistics

All the robust DOA estimation methods mentioned in the previous section assume either knowledge of the pdf or the number of mixtures. Furthermore, user-defined threshold values or weighting functions are often required. In this section we introduce techniques that are based on the sign covariance matrix (SCM) or the tau covariance matrix (TCM). These nonparametric statistics were introduced in section 3.3. The resulting estimation methods require no user-defined tuning parameters.

4.3.1 DOA estimation

Recall the signal model (2.1)

$$\mathbf{x}(n) = A(\boldsymbol{\theta})\mathbf{s}(n) + \mathbf{v}(n). \quad (4.2)$$

In this section we assume that the M -variate noise $\mathbf{v}(n)$ is from a complex spherically symmetric distribution and i.i.d. between the successive time-instants.

Assume first that the signal vector $\mathbf{s}(n)$ is i.i.d. between the successive time-instants, signals and noise are mutually independent and that the signal covariance matrix $E\{\mathbf{s}(n)\mathbf{s}^H(n)\}$ is of full rank. Denote the SCM of $\mathbf{x}(n)$ by Σ_1 and the TCM of $\mathbf{x}(n)$ by Σ_2 . Under the assumptions made, the $M - K$ smallest eigenvalues of Σ_1 are equal and the corresponding eigenvectors span the noise subspace. The same property holds for Σ_2 . Recall that M is the number of sensors and K is the number of signals. Proofs for these results are given in paper V. Moreover, the following theorem is true.

Theorem 2 *Let $\mathbf{x}(n)$ be as given in (4.2) with random $\mathbf{s}(n)$ and $\mathbf{v}(n)$. The distribution of $\mathbf{v}(n)$ is spherically symmetric, $\mathbf{v}(n)$ and $\mathbf{s}(n)$ are independent and $E\{\mathbf{s}(n)\mathbf{s}^H(n)\}$ is of rank K . Let $\mathbf{x}(1), \dots, \mathbf{x}(N)$ be a random sample from the distribution of $\mathbf{x}(n)$. Denote the sample SCM and TCM of the data by S_1 and S_2 , respectively. Let \hat{U}_1 to be the $M \times (M - K)$ matrix of eigenvectors of S_1 corresponding to the $M - K$ smallest eigenvalues. Define \hat{U}_2 to be the corresponding matrix for S_2 . Then, as $N \rightarrow \infty$,*

$$A^H \hat{U}_1 \xrightarrow{wp1} 0 \quad \text{and} \quad A^H \hat{U}_2 \xrightarrow{wp1} 0.$$

Proof. See Theorem 3 in paper V.

If the signal and noise between the successive time instants can be modeled as i.i.d., Theorem 2 implies that convergent estimates of the noise and signal subspaces can be obtained from the eigenvectors of the sample SCM or the sample TCM of the snapshot data. Consequently these estimates can be used in any subspace DOA estimation method.

The i.i.d. assumption of the signal vector can be relaxed in the case of the SCM by modeling the signal vector $\mathbf{s}(n)$ as a deterministic sequence.

Theorem 3 Let $\mathbf{x}(n) = A\mathbf{s}(n) + \mathbf{v}(n)$, $n = 1, \dots, N$ be as in (2.1) with random $\mathbf{v}(n)$ but deterministic $\mathbf{s}(n)$. Let $\mathbf{v}(1), \dots, \mathbf{v}(N)$ be i.i.d. from a spherically symmetric distribution and let $\mathbf{s}(1), \dots, \mathbf{s}(N)$ span a K -dimensional subspace. Denote the sample SCM of the data by S_1 . Then the $M - K$ smallest eigenvalues of $E\{S_1\}$ are equal and the corresponding eigenvectors are orthogonal to the columns of A . Moreover, as $N \rightarrow \infty$,

$$S_1 - E\{S_1\} \xrightarrow{wp1} 0.$$

Proof. See Theorem 4 in paper V.

The convergence of the subspace basis vectors can be shown by making an additional assumption about the signal sequence.

Theorem 4 Let $\mathbf{x}(n)$, $n = 1, \dots, N$ be as in (2.1) with $\mathbf{v}(1), \dots, \mathbf{v}(N)$ i.i.d. from a spherically symmetric distribution. Denote the sample SCM of the data by S_1 and let $\lambda_1 \geq \dots \geq \lambda_M$ be the eigenvalues of $E\{S_1\}$. Assume that there exist N_0 and $c > 0$ such that, for $N > N_0$, $\mathbf{s}(1), \dots, \mathbf{s}(N)$ span a K -dimensional subspace and $\lambda_K - \lambda_{K+1} > c$. Set \hat{U}_1 to be the $M \times (M - K)$ matrix of eigenvectors of S_1 corresponding to the $M - K$ smallest eigenvalues. Then, as $N \rightarrow \infty$,

$$A\hat{U}_1 \xrightarrow{wp1} 0.$$

Proof. See Theorem 5 in paper V.

After showing the theoretical motivation for SCM and TCM based subspace estimation methods, we are ready to give an algorithm illustrating the usage of the SCM and TCM in DOA estimation. The algorithm is presented for the SCM but the SCM can be replaced by the TCM without any additional modifications.

Algorithm SCM-TLS-ESPRIT

1. Calculate the sample SCM S_1 for the snapshots $\mathbf{x}(1), \dots, \mathbf{x}(N)$.
2. Apply the TLS-ESPRIT [94] algorithm to the eigenvectors of S_1 corresponding to the K largest eigenvalues.

An algorithm where the sample SCM is replaced by the sample TCM is called TCM-TLS-ESPRIT later in the examples.

4.3.2 Estimating the number of sources

The SCM and the TCM have also proven useful in estimating the number of sources. The following theorem has been proven in paper V.

Theorem 5 *Assume $\mathbf{x}(n)$ is distributed as given in Theorem 2. Let $\mathbf{x}(1), \dots, \mathbf{x}(N)$ be a random sample from the distribution of $\mathbf{x}(n)$. Define*

$$I_1(k) = -\log \left(\frac{\left(\prod_{i=k+1}^M \hat{\lambda}_{1,i} \right)^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^M \hat{\lambda}_{1,i}} \right)^{(M-k)N} + \frac{1}{2}k(2M-k)\log N,$$

$$I_2(k) = -\log \left(\frac{\left(\prod_{i=k+1}^M \hat{\lambda}_{2,i} \right)^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^M \hat{\lambda}_{2,i}} \right)^{(M-k)N} + \frac{1}{2}k(2M-k)\log N,$$

where $\hat{\lambda}_{1,i}$ and $\hat{\lambda}_{2,i}$, $i = k+1, \dots, M$ are the $M-k$ smallest eigenvalues of the sample SCM and TCM of the data. Let \hat{K}_1 and \hat{K}_2 be the values of k minimizing the above expressions. Then \hat{K}_1 and \hat{K}_2 are strongly consistent estimates of K .

Proof. See Theorem 6 in paper V.

Thus if we can assume that the signals and noise are i.i.d. between the successive time instants and the noise is spherically symmetric, the number of signals can be estimated by replacing the sample covariance matrix eigenvalues by the eigenvalues of the sample SCM or TCM in (2.24). Because the eigenvalues of the sample SCM or sample TCM are not estimates of the variance, whereas in the original criterion the $\hat{\lambda}_i$ are, the following technique is proposed in paper V.

1. For snapshots $X = [\mathbf{x}(1), \dots, \mathbf{x}(N)]$, calculate the sample SCM (TCM) and use its eigenvectors $\hat{\Gamma}_1$ ($\hat{\Gamma}_2$) to form the transformed observations $\hat{\Gamma}_1^H X$ ($\hat{\Gamma}_2^H X$).
2. Estimate the marginal variances of $\hat{\Gamma}_1^H X$ ($\hat{\Gamma}_2^H X$) using the sum of the squared MADs of the real and imaginary part.

3. Use the MDL criterion to estimate the number of signals from these estimates.

Note that the estimation result does not depend on the consistency correction term for the MAD.

4.3.3 Coherent signals

When some of the signals are completely coherent, the SCM and TCM can not be used in estimation of the noise or signal subspace basis vectors directly. When the array is a ULA, spatial smoothing preprocessing can be used.

The case when all the signals are completely coherent is considered in paper *IV*. If all the signal sources are completely coherent, they are phase-delayed amplitude-weighted replicas of one signal $s(n)$. In this situation, the signal vector may be given as

$$\mathbf{s}(n) = \boldsymbol{\zeta} s(n),$$

where $\boldsymbol{\zeta} = [\zeta_1, \dots, \zeta_K]^T$ is a K -vector of the complex attenuations of the signal components corresponding to the different DOAs. As in subsection 2.2.3, divide a linear uniform array with M identical sensors into overlapping forward and backward subarrays of size P ($K < P < M$). Let \mathbf{x}_l^f denote the received signals at the l th forward subarray and let \mathbf{x}_l^b denote the complex conjugate of the received signals at the l th backward subarray, $l = 1, \dots, L = M - P + 1$. See formulas (2.13)-(2.14). We can model \mathbf{x}_l^f as

$$\mathbf{x}_l^f = A D^{(l-1)} \boldsymbol{\zeta} s + \mathbf{v}_l^f$$

and

$$\mathbf{x}_l^b = A D^{(l-1)} (D^{(M-1)} \boldsymbol{\zeta} s)^* + \mathbf{v}_l^b$$

where $A = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ with $\mathbf{a}(\theta_k)$ being the $P \times 1$ ($P > K$) array steering vector corresponding to the DOA of the k th coherent signal component, D^l is given in (2.15) and \mathbf{v}_l^f and \mathbf{v}_l^b are noise vectors with spherically symmetric distribution.

Now let $\bar{\Sigma}_1$ ($\bar{\Sigma}_2$) be the average of the SCMs (TCMs) of the forward and backward subarrays and assume $M \geq 3K/2$. Paper *IV* showed that under mild restrictions on the vector $\boldsymbol{\zeta}$, it is possible to choose P such that $P - K$ smallest eigenvalues of $\bar{\Sigma}_1$ ($\bar{\Sigma}_2$) are

equal and the corresponding eigenvalues are orthogonal to the columns of the matrix A . If $L \geq K$, the restriction on $\boldsymbol{\zeta}$ is not needed. Therefore the DOAs of the coherent signals can be estimated using any subspace algorithm and an estimate of $\bar{\Sigma}_1$ or $\bar{\Sigma}_2$.

The following algorithm is proposed in paper *IV*.

Algorithm SS-SCM-MUSIC:

1. For the snapshots $\mathbf{x}(1), \dots, \mathbf{x}(N)$, calculate the sample SCM for each forward and backward subarray of size P . Denote the sample SCM of the l th forward and backward subarray by $S_{1_l}^f$ and $S_{1_l}^b$, respectively.

2. Form the matrix

$$\bar{S}_1 = \frac{\frac{1}{L} \sum_{l=1}^L S_{1_l}^f + \frac{1}{L} \sum_{l=1}^L S_{1_l}^b}{2}.$$

3. Choose the DOA estimates to be the K highest peaks in the pseudospectrum

$$V(\theta) = \frac{1}{\mathbf{a}^H(\theta) \hat{U}_1 \hat{U}_1^H \mathbf{a}(\theta)},$$

where \hat{U}_1 is the matrix of the eigenvectors of \bar{S}_1 corresponding to the $P - K$ smallest eigenvalues.

An algorithm where the SCM is replaced by the TCM is SS-TCM-MUSIC later in the examples.

The number of signals may be estimated using the following technique proposed in paper *V*.

1. For the snapshots $\mathbf{x}(1), \dots, \mathbf{x}(N)$, calculate the sample SCM for each forward and backward subarray of size P . Denote the eigenvector matrices of the sample SCM of the l th forward and backward subarrays by $\hat{\Gamma}_{1_l}^f$ and $\hat{\Gamma}_{1_l}^b$, respectively.
2. Let X_l^f and X_l^b denote the data from the l th forward and backward subarrays. Estimate the marginal variances of $\hat{\Gamma}_{1_l}^{fH} X_l^f$ by using the sum of the squared MADs of the real and imaginary part (use the consistency correction term $c = 1.4826$) and form a $P \times P$ diagonal matrix $\hat{\Lambda}_{1_l}^f$ from the obtained estimates. Do the same for the l th backward subarray to obtain a matrix $\hat{\Lambda}_{1_l}^b$.

3. Estimate the number of signals by using the eigenvalues of the matrix

$$\frac{1}{2L} \sum_{i=1}^L (\hat{\Gamma}_{1_i}^f \hat{\Lambda}_{1_i}^f \hat{\Gamma}_{1_i}^{fH} + \hat{\Gamma}_{1_i}^b \hat{\Lambda}_{1_i}^b \hat{\Gamma}_{1_i}^{bH})$$

in the modified MDL criterion given in (2.25).

Naturally the estimation can use TCMs instead of SCMs. Recall that the correction term $c = 1.4826$ ensures the MAD to be consistent when the underlying distribution is Gaussian.

4.4 Simulation results

In this section, we present various simulation results illustrating the performance of SCM and TCM based algorithms and comparing the performance to that of conventional subspace algorithms. In order to study robustness, ε -contamination and complex isotropic symmetric α -stable noise models are considered. In the sequel we use the term α -stable noise to refer to the M -variate noise distribution with i.i.d. complex isotropic symmetric α -stable margins. Note that the distribution of the resulting noise vector is spherically symmetric only when $\alpha = 2$.

In the ϵ -contaminated noise model, the M -dimensional noise is given by

$$\mathbf{v} = (1 - b)\mathbf{v}_1 + b\mathbf{v}_2 \quad (4.3)$$

where $b \sim \text{Bin}(1, \epsilon)$, $\mathbf{v}_1 \sim \mathcal{N}_C(\mathbf{0}, \sigma_1^2 I)$, $\mathbf{v}_2 \sim \mathcal{N}_C(\mathbf{0}, \sigma_2^2 I)$ (M -dimensional complex circular normal distributions). The resulting distribution is always spherically symmetric.

4.4.1 Non-coherent signals in noise

In our first simulation we use an 8 element ULA with interelement spacing equal to half a wavelength. Two independent 4-QAM communication signals of power 100 come to the array from directions $\theta_1 = 90^\circ$ and $\theta_2 = 95^\circ$. The TLS-ESPRIT, TCM-TLS-ESPRIT and SCM-TLS-ESPRIT algorithms are used to estimate the DOAs. The number of signals is assumed to be known here. The noise is modeled as α -stable noise. The values used for the characteristic exponent are $\alpha = 2$, $\alpha = 1.4$ and $\alpha = 1$. The value for the dispersion

is $\gamma = 1$ (in the Gaussian case the SNR is 14 dB). The number of snapshots used is 300. In the algorithms, the sensors indexed from 1 to $M - 1$ form the first subarray and the sensors indexed from 2 to M form the second subarray. Figure 4.1 shows histograms of the estimation results obtained from 200 Monte-Carlo realizations. In the Gaussian case, all the algorithms exhibit similar good performance. When the characteristic exponent $\alpha = 1.4$, the behavior of the conventional TLS-ESPRIT degrades and in the case of noise with extremely heavy tails ($\alpha = 1$), the TLS-ESPRIT algorithm fails to estimate the DOAs. On the other hand, the TCM-TLS-ESPRIT and SCM-TLS-ESPRIT algorithms perform reliably in the heavy-tailed noise.

Next we compare the ability of three different MDL based methods to estimate the number of signals. The methods are:

1. Standard MDL using the eigenvalues of the sample covariance matrix.
2. The estimation method introduced in 4.3.2 using the eigenvectors of the sample TCM.
3. The estimation method introduced in 4.3.2 using the eigenvectors of the sample SCM.

We use α -stable ($\gamma = 1$) and ϵ -contaminated ($\sigma_1^2 = 1$, $\sigma_2^2 = 1000$) noise distributions. The performance criterion is the relative proportion of correct estimation results and the simulation parameters are otherwise the same as in the previous simulation.

For the α -stable noise, the characteristic exponent is varied from 1 to 2 and for ϵ -contaminated noise the contamination probability is varied from 0 to 0.7. The number of independent Monte Carlo runs used is 200. The results are presented in Fig. 4.2. In the case of α -stable noise, the methods based on the SCM or TCM combined with the MAD estimate the number of signals reliably. On the other hand the conventional method fails if $\alpha < 2$. There is a difference in the robustness of the SCM- and TCM-based methods for noise from ϵ -contaminated distribution. The SCM-based method seems to be extremely robust, whereas the TCM based method is more sensitive to the contamination. The nonmonotone behavior of the curve for the standard estimation procedure is due to the fact that large contamination causes the obtained estimates to go to zero, whereas small contamination allows all possible estimation results.

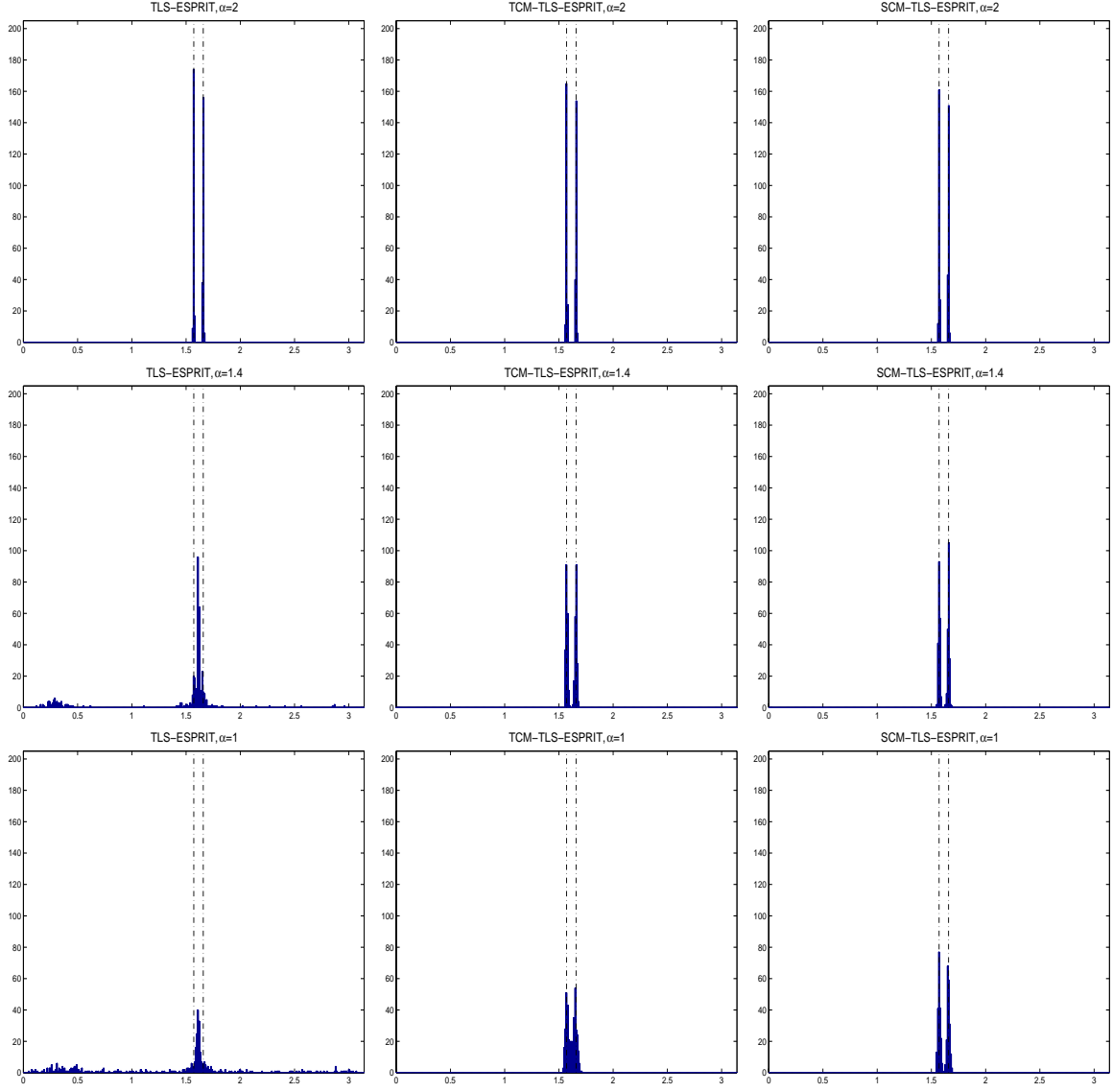


Figure 4.1: Histograms of the estimation results from 200 Monte-Carlo realizations for TLS-ESPRIT, TCM-TLS-ESPRIT and SCM-TLS-ESPRIT algorithms. The size of the ULA is 8. The DOAs are 90° and 95° .

4.4.2 Coherent signals in noise

We now compare the performance of SS-SCM-MUSIC, SS-TCM-MUSIC and forward/backward averaging with MUSIC (see [78]). We use an eight sensor ULA with interelement spacing equal to half a wavelength. The subarray size is 6. A 4-QAM communication signal of power 100 coming from 70° undergoes multipath reflection, resulting in three additional coherent arrivals from 65° , 115° and 127° . The number of signals (or

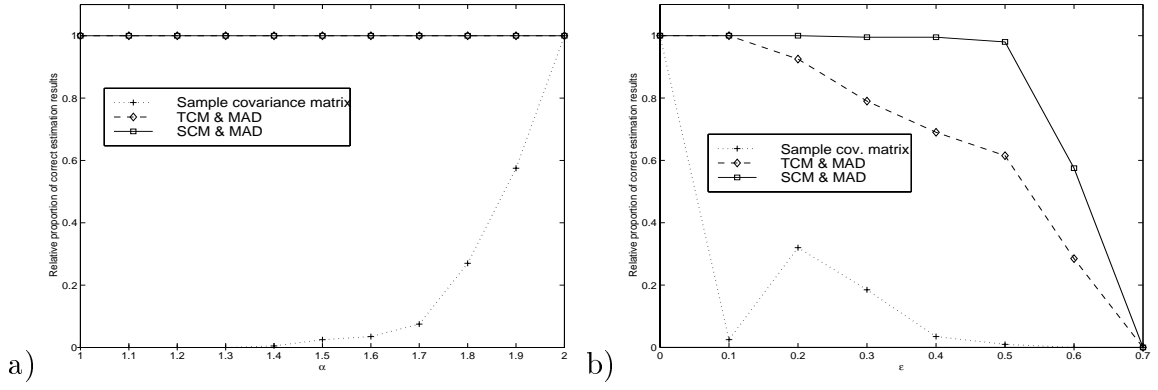


Figure 4.2: Estimating the number of signals using MDL: relative proportion of correct estimation results for a) α -stable noise as a function of α and b) ϵ -contaminated noise as a function of the contamination level ϵ .

directions) is known. The noise is α -stable noise with dispersion $\gamma = 1$. The number of snapshots is 300.

Figure 4.3 shows five estimation results for the cases $\alpha = 2$ and $\alpha = 1$. With Gaussian noise, the performance of all algorithms is almost identical. The SS-SCM-MUSIC also estimate the arrival directions well in extremely heavy-tailed noise whereas the standard spatial smoothing MUSIC algorithm fails for $\alpha = 1$. The resolution property of the SS-TCM-MUSIC does not seem to be good enough to solve the two close DOAs when $\alpha = 1$.

Next we compare the ability of three different MDL based methods to estimate the number of coherent signals. The three methods are

1. The estimation method based on a modified MDL criterion (2.25) and the eigenvalues of the forward/backward averaged sample covariance matrix.
2. The estimation method introduced in 4.3.3 using the sample TCM.
3. The estimation method introduced in 4.3.3 using the sample SCM.

We use α -stable noise with dispersion $\gamma = 1$ and ϵ -contaminated noise ($\sigma_1^2 = 1$, $\sigma_2^2 = 1000$). The simulation parameters are otherwise as in the previous simulation. For the α -stable noise, the characteristic exponent α is varied from 1 to 2 and in the case of ϵ -contaminated noise model, the contamination probability is varied from 0 to 0.7. The simulation results from 200 Monte Carlo realizations are presented in Figure 4.4. The results imply that the

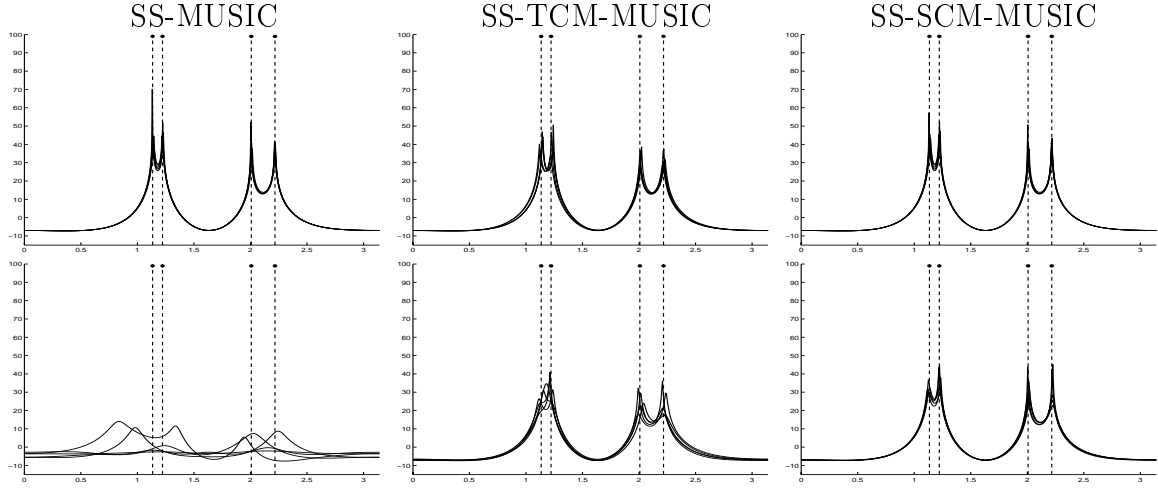


Figure 4.3: Estimating the DOAs of completely coherent sources in α -stable noise. Five results using SS-MUSIC (left column), SS-TCM-MUSIC (middle column) and SS-SCM-MUSIC (right column). First row: $\alpha = 2$. Second row: $\alpha = 1$. The size of the ULA is 8 and the subarray size is 6. The DOAs are 65° , 70° , 115° and 127° .

method combining the SCM and MAD is more robust than the method based on the TCM and MAD. The conventional method estimates the number of signals reliably only when the noise is Gaussian.

4.5 Frequency estimation using nonparametric statistics

The DOA estimation problem employing a ULA is closely related to frequency estimation of time series data. The observed time-series is given by a model of complex exponentials in white noise,

$$x(n) = \sum_{k=1}^K A_k e^{jn\omega_k} + v(n), \quad (4.4)$$

where ω_k , $k = 1, \dots, K$, are the frequencies ($\omega_l \neq \omega_p$ for $l \neq p$) and $v(n)$ is the complex-valued circular white noise. The complex amplitudes A_k are given by

$$A_k = |A_k| e^{j\phi_k},$$

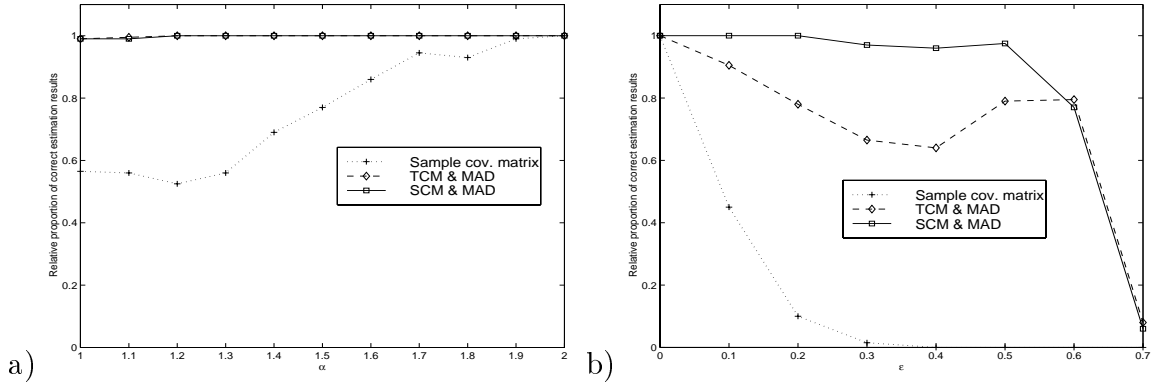


Figure 4.4: Estimating the number of coherent signals using MDL: relative proportion of correct estimation results for a) α -stable noise as a function of α and b) ϵ -contaminated noise as a function of the contamination level ϵ .

where ϕ_k is the phase. Let $\mathbf{z}(l) = [x(l), \dots, x(l + M - 1)]^T$, $M > K$, $\mathbf{a} = [A_1, \dots, A_K]^T$ and $\mathbf{v}(l) = [v(l), \dots, v(l + M - 1)]^T$. We can now write

$$\mathbf{z}(l) = BD(l)\mathbf{a} + \mathbf{v}(l),$$

where

$$B = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_K} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-1)\omega_1} & e^{j(M-1)\omega_2} & \dots & e^{j(M-1)\omega_K} \end{bmatrix}$$

and $D(l) = \text{diag}[e^{jl\omega_1}, \dots, e^{jl\omega_p}]$. It is common practice to assume that the initial phases ϕ_k are independent and uniformly distributed on $[-\pi, \pi]$, and independent of the noise. In this case the $M \times M$ autocovariance matrix of $x(n)$ is [102]

$$\Sigma = E\{\mathbf{z}(k)\mathbf{z}(k)^H\} = B\Sigma_a B^H + \sigma^2 I \quad (4.5)$$

where σ^2 is the noise variance and $\Sigma_a = \text{diag}\{|A_1|^2, \dots, |A_K|^2\}$. It then follows that the $M - K$ smallest eigenvalues of the matrix Σ are equal to the noise variance σ^2 and the corresponding eigenvectors are orthogonal to the columns of the matrix B . Therefore the subspace-methods introduced in chapter 2 can be used to estimate the frequencies ω_i . As in the case of DOA estimation, the subspace spanned by the columns of B is called signal subspace and the orthogonal subspace for it is called noise subspace.

Let $x(1), \dots, x(N)$ be an observed time series. A standard estimator for $M \times M$ autocovariance matrix is the sample autocovariance matrix

$$R = \frac{1}{N - M + 1} \sum_{i=1}^{N-M+1} \mathbf{z}(i) \mathbf{z}^H(i),$$

where $\mathbf{z}(i) = [x(i), \dots, x(i + M - 1)]^T$, $i = 1, \dots, N - M + 1$. The sample Sign Autocovariance Matrix (SAM) of size $M \times M$ is defined as

$$R_S = \frac{1}{N - M + 1} \sum_{i=1}^{N-M+1} \mathbf{S}(\mathbf{z}(i)) \mathbf{S}^H(\mathbf{z}(i)), \quad (4.6)$$

where $\mathbf{S}(\cdot)$ is the complex-valued spatial sign function defined in subsection 3.3. The following result is proven in paper VI.

Theorem 6 *Assume $\{x(1), \dots, x(N)\}$ ($N > M + K - 1$) distributed as given in (4.4) with deterministic or stochastic initial phases $\boldsymbol{\phi} = [\phi_1, \dots, \phi_K]^T$ and assume the noise $v(n)$ to be i.i.d. circular Gaussian and independent of the phases. Denote the $M \times M$ ($M > K$) sample SAM of the data by R_S . Then*

(i) *The $M - K$ smallest eigenvalues of $E\{R_S|\boldsymbol{\phi}\}$ are equal and the corresponding eigenvectors are orthogonal to the columns of the matrix B .*

(ii) *As $N \rightarrow \infty$,*

$$R_S - E\{R_S|\boldsymbol{\phi}\} \xrightarrow{w.p.1} 0.$$

Proof. See Theorem 1 in paper VI¹.

This theorem shows that the SAM can be used to find convergent estimates of the signal and noise subspace basis vectors, if the noise is assumed to be Gaussian. To illustrate the efficiency and robust performance of the SAM-based subspace methods in non-Gaussian noise, we perform a small simulation study. We compare the performance of the TLS-ESPRIT algorithm to that of the following algorithm proposed in paper VI.

¹In Theorem 1 of paper VI, $N > p$ should be $N > M + p - 1$ and $E\{R_S\}$ should be $E\{R_S|\boldsymbol{\phi}\}$

Algorithm SAM-TLS-ESPRIT

1. Calculate R_S of the size $M \times M$ for the data $x(1), \dots, x(N)$. Set \hat{U}_s to be the $M \times K$ matrix of the eigenvectors of R_S corresponding to the K largest eigenvalues.

2. Calculate the total least squares estimate $\hat{\Psi}$ for

$$J_1 \hat{U}_s \hat{\Psi} \approx J_2 \hat{U}_s,$$

where $J_1 = [I_{M-1} \ \mathbf{0}]$ and $J_2 = [\mathbf{0} \ I_{M-1}]$.

3. The frequency estimates are $\hat{\omega}_k = -j \arg(\hat{\nu}_k)$, where $\hat{\nu}_k$, $k = 1 \dots, K$ are the eigenvalues of $\hat{\Psi}$.

We use the following signal model

$$x(n) = \sum_{i=1}^4 \sqrt{50} e^{j\omega_i n + \phi_i} + v(n),$$

where $\omega_1 = 91/72\pi$, $\omega_2 = 89/72\pi$, $\omega_3 = 14/18\pi$, $\omega_4 = 13/18\pi$ and ϕ_i , $i = 1, \dots, 4$ are uniformly distributed on $[-\pi, \pi]$. The number of observations is $N = 300$ and the size of the matrices (number of lags) used in the algorithms is $M = 30$. The noise is generated from complex isotropic symmetric α -stable noise distribution with dispersion $\gamma = 1$. The values used for the characteristic exponent are $\alpha = 1$ and $\alpha = 2$.

Figure 4.5 shows histograms for the estimation results obtained from 200 Monte-Carlo realizations using TLS-ESPRIT and SAM-TLS-ESPRIT algorithms. When the noise is Gaussian, the performance of the two methods is almost identical. When the noise is non-Gaussian ($\alpha = 1$), the SAM-TLS-ESPRIT estimates the frequencies significantly better than the TLS-ESPRIT.

Note that the number of complex exponentials K is known in the simulations. If K is not known *a priori*, it has to be estimated from the data. Common practice is to use MDL criterion for this task. In paper VI a robust MDL-based method is proposed for estimation of the number of complex exponentials, and its robust performance is verified in simulation.

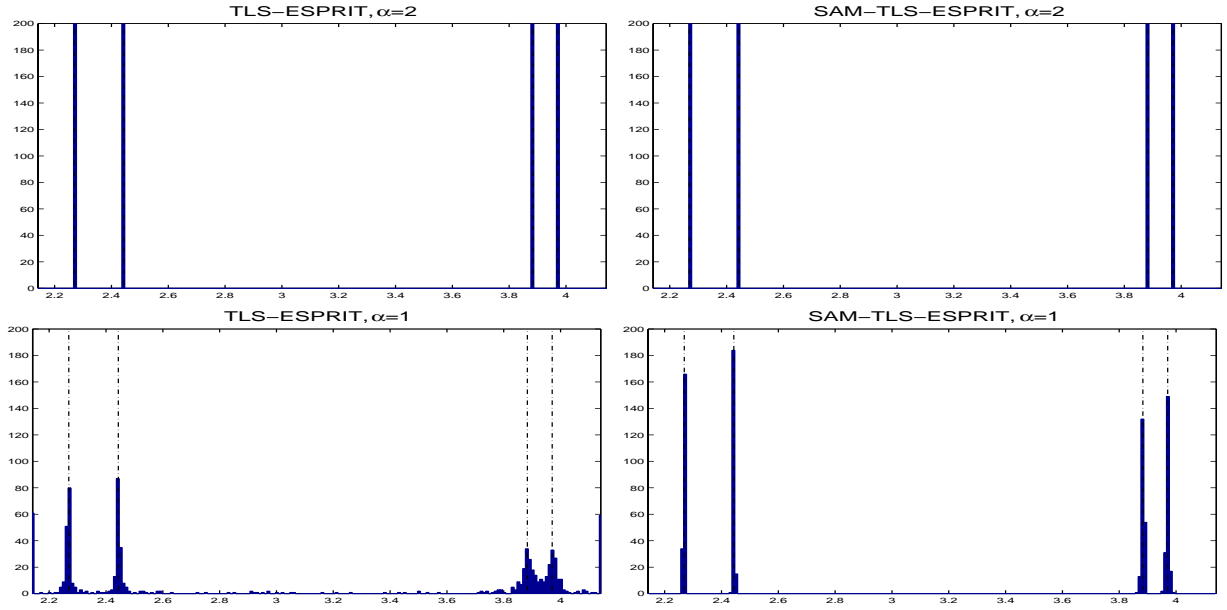


Figure 4.5: Histograms of the estimation results from 200 Monte-Carlo realizations for TLS-ESPRIT and SAM-TLS-ESPRIT algorithms. The number of observations is $N = 300$ and the number of lags is $M = 30$.

4.6 Discussion

In this chapter we presented the main contributions of the thesis. It was shown that convergent estimates of signal and noise subspace basis vectors can be obtained from the sample SCM or the sample TCM. Algorithms were proposed based on these nonparametric statistics. Moreover, an extension to frequency estimation was given. The resulting estimation techniques require no user defined tuning parameters. Therefore they are easier to apply successfully than the other robust estimation methods appearing in the literature.

The simulation results show that DOA estimation algorithms based on nonparametric statistics have reliable performance and closely spaced sources can be resolved regardless of the heavy-tailed nature of the noise distribution. Also the robust techniques for estimation of the number of signals have good performance in Gaussian and non-Gaussian noise. In general, the behavior (resolution, number of snapshots required, the effect of SNR) of these algorithms is very similar to the standard algorithms when the noise is Gaussian. When the noise is non-Gaussian, the proposed robust algorithms perform consistently better.

Chapter 5

Summary

Antenna array signal processing has a remarkable role in future wireless communication systems. Other application areas include radar, sonar and biomedicine. In most array signal processing applications the first task is to estimate the number of source signals and their DOAs. The conventional algorithms for these tasks rely heavily on the sample covariance matrix, and can fail to perform reliably if the noise appearing in the measurements are non-Gaussian. Real-world measurement noise may deviate significantly from the Gaussian assumption. Consequently robust estimation methods that have reliable albeit sub-optimal performance in both Gaussian and non-Gaussian noise have been developed.

The robust estimation methods proposed in this thesis are based on multivariate non-parametric statistics. The techniques are simple and require no user-defined tuning parameters. Therefore they are easier to apply successfully than the other robust methods appearing in the literature. In the thesis it is proven that the sample SCM and the sample TCM can be applied to obtain convergent estimates of the signal and noise subspaces basis vectors. These estimates may then be used in any subspace-based DOA estimation method, such as MUSIC or ESPRIT. When the array is a ULA, spatial smoothing pre-processing can be used to deal with coherent signals. In the proofs it is assumed that the noise has a complex spherically symmetric distribution. The family of complex spherically symmetric distributions is large and includes spatially white circular complex Gaussian distributions. The techniques proposed for estimation of the number of source signals are

based on combined use of the sample SCM or TCM and a robust estimator of variance. The use of nonparametric statistics is also proposed in frequency estimation and the theoretical motivation of the resulting techniques is proven. Simulation results show that the proposed methods have reliable performance in Gaussian and non-Gaussian noise.

A large number of signal processing applications other than array processing require processing of multichannel data and estimation of the covariance matrix. If the data is non-Gaussian, robust covariance matrix estimators should be used. In the thesis it is proven that the eigenvectors of the SCM and the TCM are the same as the eigenvectors of the ordinary covariance matrix for a large family of symmetric distributions. The sample SCM and TCM may therefore be used to construct robust estimates of the covariance matrix with a combined use of robust univariate scale estimator such as the MAD. Numerical calculation of these estimates is straightforward which is a remarkable benefit when compared to other robust covariance matrix estimation methods. The weakness of the proposed approach is the lack of affine equivariance.

Future research continuing the work of this thesis might include the use of the SCM and the TCM in applications other than array signal processing where the low rank signal model (2.1) is valid. These applications include blind channel estimation, and delay estimation in communication systems. Further investigation of the asymptotic properties of the proposed DOA estimation algorithms would be useful. To this end asymptotic distributions of the eigenvalues and eigenvectors of the sample SCM should be derived. The location scale family considered in the case of covariance matrix estimation does not generally cover the low rank signal model used in array signal processing. However when the signal and noise are complex circular Gaussian or the joint distribution of the signal and noise is complex elliptically symmetric the model is appropriate for the distribution of the array output vector. Therefore the covariance matrix estimates based on the sample SCM or the sample TCM may also be used in the SML or the WSF method. The performance of this kind of combined methods is an interesting issue to be studied.

Appendix A

Proof of Theorem 1

Theorem 1 *Let \mathbf{z} be an M -variate random variable with a reflection and permutation invariant distribution. Consider a random variable*

$$\mathbf{x} = U D \mathbf{z}$$

where $D = \text{diag}\{d_1, \dots, d_M\}$, $|d_1| \geq \dots \geq |d_M| > 0$ and U is a unitary matrix. Let the covariance matrix and SCM of \mathbf{x} be Σ and Σ_1 , respectively. Then

$$\Sigma = U \Lambda U^H \quad \text{and} \quad \Sigma_1 = U \Lambda_1 U^H$$

where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_M\}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > 0$, and $\Lambda_1 = \text{diag}\{\lambda_{1,1}, \dots, \lambda_{1,M}\}$, $\lambda_{1,1} \geq \lambda_{1,2} \geq \dots \geq \lambda_{1,M} > 0$. Moreover $\lambda_i = \lambda_{i+1}$ if and only if $\lambda_{1,i} = \lambda_{1,i+1}$, or in other words the eigenvectors of Σ and Σ_1 ordered by their respective eigenvalues can be chosen to be the same.

Proof. We assume for simplicity that $P(\mathbf{z}^H \mathbf{z} = 0) = 0$. If it were $0 < P(\mathbf{z}^H \mathbf{z} = 0) < 1$, the expectations below are understood as conditional on $\mathbf{z} \neq \mathbf{0}$. Because the distribution of \mathbf{z} is reflection invariant,

$$\begin{aligned} E\{\mathbf{S}(D\mathbf{z})\mathbf{S}^H(D\mathbf{z})\} &= E\{\mathbf{S}(DG\mathbf{z})\mathbf{S}^H(DG\mathbf{z})\} \\ &= GE\{\mathbf{S}(D\mathbf{z})\mathbf{S}^H(D\mathbf{z})\}G^T \end{aligned}$$

for any reflection matrix G . Therefore $E\{\mathbf{S}(D\mathbf{z})\mathbf{S}^H(D\mathbf{z})\} = \Lambda_1 = \text{diag}\{\lambda_{1,1}, \dots, \lambda_{1,M}\}$, where

$$\begin{aligned}\lambda_{1,i} &= E \left\{ \frac{|d_i|^2 |z_i|^2}{\sum_{k=1}^M |d_k|^2 |z_k|^2} \right\} \\ &= E \left\{ \frac{|z_i|^2}{\sum_k (|d_k|^2 / |d_i|^2) |z_k|^2} \right\}, \quad i = 1, \dots, M.\end{aligned}$$

Fix $i < M$ and denote by \sum' the summation over k excluding $k = i, i+1$. Firstly, by permutation invariance

$$\begin{aligned}\lambda_{1,i} &= E \left\{ \frac{|z_i|^2}{|z_i|^2 + (|d_{i+1}|^2 / |d_i|^2) |z_{i+1}|^2 + \sum' (|d_k|^2 / |d_i|^2) |z_k|^2} \right\} \\ &= E \left\{ \frac{|z_{i+1}|^2}{|z_{i+1}|^2 + (|d_{i+1}|^2 / |d_i|^2) |z_i|^2 + \sum' (|d_k|^2 / |d_i|^2) |z_k|^2} \right\}.\end{aligned}$$

Then, because $|d_{i+1}|/|d_i| \leq 1 \leq |d_i|/|d_{i+1}|$ and $|d_k|/|d_i| \leq |d_k|/|d_{i+1}|$ for all k , we find

$$\begin{aligned}& E \left\{ \frac{|z_{i+1}|^2}{|z_{i+1}|^2 + (|d_{i+1}|^2 / |d_i|^2) |z_i|^2 + \sum' (|d_k|^2 / |d_i|^2) |z_k|^2} \right\} \\ & \geq E \left\{ \frac{|z_{i+1}|^2}{|z_{i+1}|^2 + (|d_i|^2 / |d_{i+1}|^2) |z_i|^2 + \sum' (|d_k|^2 / |d_{i+1}|^2) |z_k|^2} \right\} \\ & = E \left\{ \frac{|d_{i+1}|^2 |z_{i+1}|^2}{|d_{i+1}|^2 |z_{i+1}|^2 + |d_i|^2 |z_i|^2 + \sum' |d_k|^2 |z_k|^2} \right\} \\ & = \lambda_{1,i+1}.\end{aligned}$$

Thus $\lambda_{1,1} \geq \dots \geq \lambda_{1,M}$. It is clear that $\lambda_{1,i} = \lambda_{1,i+1}$ if and only if $|d_i| = |d_{i+1}|$. Let U be an arbitrary unitary matrix. Then

$$E\{\mathbf{S}(UD\mathbf{z})\mathbf{S}^H(UD\mathbf{z})\} = U\Lambda_1 U^H.$$

By reflection and permutation invariance $E(\mathbf{z}\mathbf{z}^H) = wI$. Then the ordinary covariance matrix is

$$E\{UD\mathbf{z}\mathbf{z}^H D^H U^H\} = U\Lambda U^H,$$

where $\Lambda = w D^H D = w \text{diag}\{|d_1|^2, \dots, |d_M|^2\}$ with $w = E(|z_i|^2)$. Therefore the same unitary matrix diagonalizes both the ordinary covariance matrix Σ and the sign covariance matrix Σ_1 of $\mathbf{x} = UD\mathbf{z}$. Moreover, if $U^H \Sigma U = \text{diag}\{\lambda_1, \dots, \lambda_M\}$ satisfy $\lambda_1 \geq \dots \geq \lambda_M$ then also $\lambda_{1,1} \geq \dots \geq \lambda_{1,M}$ from $\text{diag}\{\lambda_{1,1}, \dots, \lambda_{1,M}\} = U^H \Sigma_1 U$. Note that $\lambda_i = \lambda_{i+1}$ if and only if $\lambda_{1,i} = \lambda_{1,i+1}$.

Bibliography

- [1] J. Agulló. Exact iterative computation of the multivariate minimum volume ellipsoid estimator with a branch and bound algorithm. In A. Prat, editor, *Proc. in Computational Statistics*, pages 175–180. Physica-Verlag, 1996.
- [2] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B.N Petrov and F. Caski, editors, *Proc. of the 2nd International Symposium on Information Theory*, pages 267–281, 1973.
- [3] K. S. Arun and B. D. Rao. An improved Toeplitz approximation method. In *Proc. of the 1988 IEEE Int'l Conf. on Acoust. Speech and Signal Proc.*, pages 2352–2355, 1988.
- [4] A. J. Barabell. Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *Proc. of the 1983 IEEE Int'l Conf. on Acoust. Speech and Signal Proc.*, pages 336–339, 1983.
- [5] H. Bensmail and G. Celeux. Regularized Gaussian discriminant analysis through eigenvalue decomposition. *Journal of the American Statistical Association*, 91:1743–1749, 1996.
- [6] J.F. Böhme. Estimation of spectral parameters of correlated signals in wavefields. *Signal Processing*, 10:329–337, 1986.
- [7] J.F. Böhme and D. Kraus. On least squares methods for direction of arrival estimation in the presence of unknown noise fields. In *Proc. of the 1988 IEEE Int'l Conf. on Acoust. Speech and Signal Proc.*, pages 2833–2836, 1988.

- [8] Y. Bresler. Maximum likelihood estimation of linearly structured covariance with application to antenna array processing. In *Proc. of the 4th ASSP Workshop on Spectrum Estimation and Modeling*, pages 172–175, 1998.
- [9] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics Speech and Signal Processing*, 34:1081–1088, 1986.
- [10] K. Buckley and X.-L. Xu. Spatial-spectrum estimation in a location sector. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38(11):1842–1852, 1990.
- [11] R. W. Butler, P. L. Davies, and M. Jhun. Asymptotics for the minimum covariance determinant estimator. *The Annals of Statistics*, 21:1385–1400, 1993.
- [12] J. A. Cadzow. A high resolution direction-of-arrival algorithm for narrow-band coherent and incoherent sources. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(7):965–979, 1988.
- [13] N. A. Campbell. Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, 29(3):231–237, 1980.
- [14] N. A. Campbell, H. P. Lopuhaä, and P. J. Rousseeuw. On the calculation of a robust S-estimator of a covariance matrix. *Statistics in Medicine*, 17:2685–2695, 1998.
- [15] J. Capon. High resolution frequency–wavenumber spectral analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.
- [16] P. Chaudhuri. Multivariate location estimation using extension of R -estimates through U -statistics type approach. *The Annals of Statistics*, 20:897–916, 1992.
- [17] P. Chaudhuri. On a geometrical notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91:862–872, 1996.
- [18] H. H. Chiang and C. L. Nikias. The ESPRIT algorithm with higher-order statistics. In *Proc. of Workshop on Higher-Order Spectral Analysis*, pages 163–168, 1989.

- [19] R. D. Cook, D. M. Hawkins, and S. Weisberg. Exact iterative computation of the robust multivariate minimum volume ellipsoid estimator. *Statistics & Probability Letters*, 16:213–218, 1993.
- [20] P. Costa, J. Grouffaud, P. Larzabal, and Henre Clergeot. Estimation of the number of signals from features of the covariance matrix: A supervised approach. *IEEE Transactions on Signal Processing*, 47(11):3108–3115, 1999.
- [21] C. Croux and G. Haesbroeck. An easy way to increase the finite-sample efficiency of the resampled minimum volume ellipsoid estimator. *Computational Statistics and Data Analysis*, 25:125–141, 1997.
- [22] C. Croux and G. Haesbroeck. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, 71(2):161–190, 1999.
- [23] C. Croux and G. Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Preprint, University of Brussels*, 2000.
- [24] C. Croux and A. Ruiz-Gazen. A fast algorithm for robust principal components based on projection pursuit. In A. Prat, editor, *Compstat: Proceedings in Computational Statistics*, pages 211–217. Physica-Verlag, 1996.
- [25] L. Davies. The asymptotics of Rousseeuw’s minimum volume ellipsoid estimator. *The Annals of Statistics*, 20(4):1828–1843, 1992.
- [26] M. C. Dogan and J. M. Mendel. Application of cumulants to array signal processing—part I: Aperture extension and array calibration. *IEEE Transactions on Signal Processing*, 43(5):1200–1216, 1995.
- [27] D. L. Donoho. Breakdown properties of multivariate location estimators. Ph.D. qualifying paper. Harvard University, 1982.

- [28] D. L. Donoho and P. J. Huber. The notion of breakdown point. In P. Bickel, K. Doksum, and J. L. Hodges Jr., editors, *A Festschrift for Erich Lehmann*, pages 157–184, 1983.
- [29] K.-T. Fang, S. Kotz, and K. W. Ng. *Symmetric multivariate and related distributions*. Chapman and Hall, London, 1990.
- [30] K.-T. Fang and Y.-T. Zhang. *Generalized Multivariate Analysis*. Springer-Verlag, 1990.
- [31] M. Feder. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(4):477–489, 1988.
- [32] J. A. Fessler and A. O. Hero. Space-alternating generalized EM algorithm. *IEEE Transactions on Signal Processing*, 42(10):2664–2677, 1995.
- [33] M. Haardt and J. A. Nossék. Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden. *IEEE Transactions on Signal Processing*, 43(5):1232–1242, 1995.
- [34] F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42:1887–1896, 1971.
- [35] F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69:383–393, 1974.
- [36] F.R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986.
- [37] D. A. Harville. *Matrix Algebra From a Statistician’s Perspective*. Springer-Verlag, 1997.
- [38] D. M. Hawkins. A feasible solution algorithm for the minimum volume ellipsoid estimator in multivariate data. *Computational Statistics*, 9:95–107, 1993.

- [39] D. M. Hawkins. The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data. *Computational Statistics and Data Analysis*, 17:197–210, 1994.
- [40] J. L. Hodges Jr. Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 163–168, 1967.
- [41] Y. Hua. The most efficient implementation of IQML algorithm. *IEEE Transactions on Signal Processing*, 42(8):2203–2204, 1994.
- [42] P. J. Huber. *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics, 1977.
- [43] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.
- [44] A. G. Jaffer. Maximum likelihood direction finding of stochastic sources: A separable solution. In *Proc. of the 1988 IEEE Int’l Conf. on Acoust. Speech and Signal Proc.*, volume 5, pages 2893–2896, 1988.
- [45] N. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, 1972.
- [46] J. P. F. J. Juan. A subsampling method for the computation of multivariate estimators with high breakdown point. *Journal of Computational and Graphical Statistics*, 3(4):319–334, 1995.
- [47] S. A. Kassam and H. V. Poor. Robust techniques for signal processing: a survey. *Proceedings of the IEEE*, 73(3):433–481, 1985.
- [48] M. Kaveh and A. J. Barabell. The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise. *IEEE Transactions on Acoustics Speech and Signal Processing*, 34(2):331–341, 1986.
- [49] J. T. Kent and D. E. Tyler. Redescending M-estimates for multivariate location and scatter. *The Annals of Statistics*, 19:2102–2119, 1991.

- [50] J. T. Kent and D. E. Tyler. Constrained M-estimation for multivariate location and scatter. *The Annals of Statistics*, 24:1346–1370, 1996.
- [51] V. Koivunen. Nonlinear filtering of multivariate images under robust error criterion. *IEEE Transactions on Image Processing*, 5(6):1054–1060, 1996.
- [52] R. J. Kozick. *Signal Processing for High-resolution Array Imaging and Adaptive Beamforming*. PhD thesis, University of Pennsylvania, 1992.
- [53] R. J. Kozick and S. A. Kassam. A unified approach to coherent source decorrelation by autocorrelation matrix smoothing. *Signal Processing*, 45(1), 1995.
- [54] R. J. Kozick and B. M. Sadler. Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures. *IEEE Transactions on Signal Processing*, 48(12), 2000.
- [55] H. Krim and M. Viberg. Two decades of array signal processing research: The parametric approach. *IEEE Signal Processing Magazine*, 13(4):67–94, 1996.
- [56] P.R. Krishnaiah and J. Lin. Complex elliptically symmetric distributions. *Communications in Statistics, Part A – Theory and Methods*, 15(12):3693–3718, 1986.
- [57] R. Kumaresan and D. W. Tufts. Estimating the angles of arrival of multiple plane waves. *Transactions on Aerospace and Electronic Systems*, 19:134–139, 1983.
- [58] S. Y. Kung, C.K. Lo, and R. Foka. A Toeplitz approximation approach to coherent source direction finding. In *Proc. of the 1986 IEEE Int’l Conf. on Acoust. Speech and Signal Proc.*, pages 193–196, 1986.
- [59] D. D. Lee and R. L. Kashyap. Robust maximum likelihood bearing estimation in contaminated Gaussian noise. *IEEE Transactions on Signal Processing*, 40(8):1983–1986, 1992.
- [60] D. D. Lee, R. L. Kashyap, and R. N. Madan. Robust decentralized direction-of-arrival estimation in contaminated noise. *IEEE Transactions on Signal Processing*, 38(3):496–505, 1990.

- [61] G. Li and Z. Chen. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and monte-carlo. *Journal of the American Statistical Association*, 80:759–766, 1985.
- [62] H. P. Lopuhaä. On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*, 17(4), 1989.
- [63] H. P. Lopuhaä and P. J. Rousseeuw. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, 19(1):229–248, 1991.
- [64] J. I. Marden. Some robust estimates of principal components. *Statistics & Probability Letters*, 43(4):349–359, 1999.
- [65] R. A. Maronna, W. A. Stahel, and V. J. Yohai. Bias-robust estimators of multivariate scatter based on projections. *Journal of Multivariate Analysis*, 42:141–161, 1992.
- [66] R. A. Maronna and V. J. Yohai. The behavior of the Stahel-Donoho multivariate estimator. *Journal of the American Statistical Association*, 90:330–341, 1995.
- [67] R. A. Maronna and V. J. Yohai. Robust estimation of multivariate location and scatter. In S. Kotz, C. Read, and D. Banks, editors, *Encyclopedia of Statistical Sciences*, update volume 2, 1998.
- [68] R.A. Maronna. Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, 20, 1976.
- [69] D. Middleton. Man-made noise in urban environments and transportation systems: Models and measurements. *IEEE Transactions on Communications*, 21:1232–1241, 1973.
- [70] M.I. Miller and D.R. Fuhrmann. Maximum-likelihood narrow-band direction finding and the EM algorithm. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38(9):1560–1577, 1990.

- [71] J. Möttönen and H. Oja. Multivariate spatial sign and rank methods. *Nonparametric Statistics*, 5:201–213, 1995.
- [72] A. Nehorai, D. Starer, and P. Stoica. Direction-of-arrival estimation in applications with multipath and few snapshots. *Circuits, Systems, and Signal Processing*, 10:327–342, 1991.
- [73] B. Ottersten, M. Viberg, and T. Kailath. Performance analysis of the total least squares ESPRIT algorithm. *IEEE Transactions on Signal Processing*, 39(5):1122–1135, 1991.
- [74] B. Ottersten, M. Viberg, and T. Kailath. Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Transactions on Signal Processing*, 40(3):590–600, 1992.
- [75] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. Exact and large sample ML techniques for parameter estimation and detection in array processing. In Haykin, Litva, and Shepherd, editors, *Radar Array Processing*, pages 99–151. Springer-Verlag, 1993.
- [76] A. Paulraj, R. Roy, and T. Kailath. A subspace rotation approach to signal parameter estimation. *Proceedings of the IEEE*, pages 1044–1045, 1986.
- [77] S. U. Pillai. *Array Signal Processing*. Springer-Verlag, New York, 1989.
- [78] S. U. Pillai and H.K. Byung. Forward/backward spatial smoothing techniques for coherent signal identification. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(1):8–15, 1989.
- [79] S. U. Pillai and B.H. Kwong. Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(8):1176–1189, 1989.
- [80] V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astron. Soc.*, 33:347–366, 1973.

- [81] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 2 edition, 1994.
- [82] B. Porat and B. Friedlander. Analysis of the asymptotic relative efficiency of the MUSIC algorithm. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(4):532–543, 1988.
- [83] B. D. Rao and K. V. S. Hari. Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(12):1990–1995, 1989.
- [84] B. D. Rao and K. V. S. Hari. Effect of spatial smoothing on the performance of MUSIC and the minimum-norm method. *IEE Proceedings*, 137(6):449–458, 1990.
- [85] B. D. Rao and K. V. S. Hari. Weighted subspace methods and spatial smoothing: analysis and comparison. *IEEE Transactions on Signal Processing*, 41(2):788–803, 1993.
- [86] B.D. Rao and K.V.S Hari. Effect of spatial smoothing on the performance of noise subspace based methods. In *Proc. of the 1990 IEEE Int'l Conf. on Acoust. Speech and Signal Proc.*, pages 2687–2690, 1990.
- [87] B.D. Rao and K.V.S Hari. Effect of spatial smoothing on state space methods/ESPRIT. In *Proc. of the 1991 IEEE Int'l Conf. on Acoust. Speech and Signal Proc.*, 1991.
- [88] J. Rissanen. Modeling by the shortest data description. *Automatica*, 14:465–471, 1978.
- [89] P. Rousseeuw. Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, editors, *Proc. of the 4th Panonian Symp. on Math. Stat.*, pages 283–297, 1983.

- [90] P. J. Rousseeuw and S. Van Aelst. Positive-breakdown robust methods in computer vision. In K. Berk and M. Pourahmadi, editors, *Computing Science and Statistics*, volume 31, pages 451–460, 1999.
- [91] P. J. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.
- [92] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.
- [93] P. J. Rousseeuw and B. C. van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411):633–639, 1990.
- [94] R. Roy and T. Kailath. ESPRIT– estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(7):984–995, 1989.
- [95] D. Ruppert. Computing S estimators for regression and multivariate location/dispersion. *Journal of Computational and Graphical Statistics*, 1(3):253–270, 1992.
- [96] R. Schmidt. Multiple emitter location and signal parameter estimation. In *Proc. of the RADC Spectrum Estimation Workshop*, pages 243–258, 1979.
- [97] M. Shao and C. L. Nikias. *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, New-York, 1995.
- [98] Splus. MathSoft, 1997.
- [99] W. A. Stahel. Breakdown of covariance estimators. *Research Report 31, Fachgruppe für Statistik*, E.T.H. Zurich, 1981.
- [100] D. Starer and A. Nehorai. Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise. *IEEE Transactions on Signal Processing*, 40(6):1528–1534, 1992.

- [101] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.
- [102] P. Stoica and R. L. Moses. *Introduction to Spectral Analysis*. Wiley, New York, 1997.
- [103] P. Stoica and A. Nehorai. MUSIC, maximum likelihood and Cramér-Rao bound. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(5):720–741, 1989.
- [104] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38(10):1783–1795, 1990.
- [105] P. Stoica and A. Nehorai. MUSIC, maximum likelihood and Cramér-Rao bound: Further results and comparisons. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38(12):2140–2150, 1990.
- [106] P. Stoica and A. Nehorai. Performance comparison of subspace rotation and MUSIC methods for direction estimation. *IEEE Transactions on Signal Processing*, 39(2):446–453, 1991.
- [107] P. Stoica, B. Ottersten, M. Viberg, and R.L. Moses. Maximum likelihood array processing for stochastic coherent sources. *IEEE Transactions on Signal Processing*, 44(1):96 – 105, 1996.
- [108] A. L. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath. Multiple invariance ESPRIT. *IEEE Transactions on Signal Processing*, 40(4):867–881, 1992.
- [109] J. S. Thompson, P. M. Grant, and B. Mulgrew. Performance of spatial smoothing algorithms for correlated sources. *IEEE Transactions on Signal Processing*, 44(4):1040–1046, 1996.
- [110] P. Tsakalides and C. L. Nikias. Maximum likelihood localization of sources in noise modeled as a stable process. *IEEE Transactions on Signal Processing*, 43(11):2700–2713, 1995.

- [111] P. Tsakalides and C. L. Nikias. The robust covariation-based music (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments. *IEEE Transactions on Signal Processing*, 44(7):1623–1633, 1996.
- [112] D. E. Tyler. Some results on the existence, uniqueness, and computation of the m -estimates of multivariate location and scatter. *SIAM Journal on Scientific and Statistical Computing*, 9:354–362, 1988.
- [113] D. E. Tyler. Finite sample breakdown points of projection based multivariate location and scatter statistics. *The Annals of Statistics*, 22:1024–1044, 1994.
- [114] M. Viberg, B. Ottersten, and T. Kailath. Detection and estimation in sensor arrays using weighted subspace fitting. *IEEE Transactions on Signal Processing*, 39(11):2436–2449, 1991.
- [115] X. Wang and H. V. Poor. Robust adaptive array for wireless communications. *IEEE Transactions on Signal Processing*, 16(8):1352–1366, 1998.
- [116] M. Wax. Detection and localization of multiple sources via the stochastic signals model. *IEEE Transactions on Signal Processing*, 39(11):2450–2456, 1991.
- [117] M. Wax. Detection and localization of multiple sources in noise with unknown covariance. *IEEE Transactions on Signal Processing*, 30(1):245–249, 1992.
- [118] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33:387–392, 1985.
- [119] M. Wax and I. Ziskind. Detection of the number of coherent signals by the mdl principle. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(8):1190–1199, 1989.
- [120] M. Wax and I. Ziskind. On unique localization of multiple sources by passive sensor arrays. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(7):996–1000, 1989.

- [121] D. B. Williams and D. H. Johnson. Robust estimation of structured covariance matrices. *IEEE Transactions on Signal Processing*, 41(9):2891–2906, 1994.
- [122] D. L. Woodruff and D. M. Rocke. Computable robust estimation of multivariate location and shape in high dimensions using compound estimators. *Journal of the American Statistical Association*, 89(427):888–896, 1994.
- [123] H.-T. Wu, J.-F. Yang, and F.-K. Chen. Source number estimators using transformed gerschgorin radii. *IEEE Transactions on Signal Processing*, 43(6):1325–1333, 1995.
- [124] G. Xu, R. H. Roy, and T. Kailath. Detection of number of sources via exploitation of centro-symmetry property. *IEEE Transactions on Signal Processing*, 42(1), 1994.
- [125] X-L Xu and K. Buckley. Bias analysis of the MUSIC location estimator. *IEEE Transactions on Signal Processing*, 40(10):2559–2569, 1992.
- [126] Y. Yardimci, A. E. Çetin, and J. A. Cadzow. Robust direction-of-arrival estimation in non-Gaussian noise. *IEEE Transactions on Signal Processing*, 46(5):1443–1451, 1998.
- [127] N. Yuen and B. Friedlander. Asymptotic performance analysis of ESPRIT, higher order ESPRIT, and virtual ESPRIT algorithms. *IEEE Transactions on Signal Processing*, 44(10):2537–2550, 1996.
- [128] Q. T. Zhang. Probability of resolution of the MUSIC algorithm. *IEEE Transactions on Signal Processing*, 43(4):978–987, 1995.
- [129] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals in presence of white noise. *Journal of Multivariate Analysis*, 20(1):1–25, 1986.
- [130] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals when the noise covariance matrix is arbitrary. *Journal of Multivariate Analysis*, 20:26–49, 1986.

- [131] Z. Zhu, S. Haykin, and X. Huang. Estimating the number of signals using reference noise samples. *IEEE Transactions on Aerospace and Electronic Systems*, 27(3):575–578, 1991.
- [132] I. Ziskind and M. Wax. Maximum likelihood localization of multiple sources by alternating projection. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(10):1553–1560, 1988.