

WIDE - BAND MAXIMUM LIKELIHOOD DIRECTION
FINDING BY USING THE TREE - STRUCTURED EM
ALGORITHM

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Hail Çadallı
July 1996

QA
276.8
.C33
1996

WIDE-BAND MAXIMUM LIKELIHOOD DIRECTION
FINDING BY USING THE TREE-STRUCTURED EM
ALGORITHM

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By

Nail Çadallı

July 1996

NAIL ÇADALLI

tarafından beğenilmiştir.

QA

276.8

. C33

1998

B. 034655

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



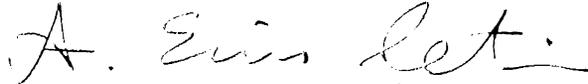
Asst. Prof. Dr. Orhan Arıkan(Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



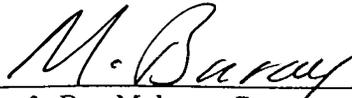
Asst. Prof. Dr. Billur Barshan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Assoc. Prof. Dr. A. Enis Çetin

Approved for the Institute of Engineering and Sciences:



Prof. Dr. Mehmet Baray
Director of Institute of Engineering and Sciences

ABSTRACT

WIDE-BAND MAXIMUM LIKELIHOOD DIRECTION FINDING BY USING THE TREE-STRUCTURED EM ALGORITHM

Nail Çadallı

M.S. in Electrical and Electronics Engineering

Supervisor: Asst. Prof. Dr. Orhan Arıkan

July 1996

A thorough derivation of the Expectation Maximization (EM) algorithm, which is an iterative numerical method of Maximum Likelihood (ML) estimation, is presented for the case of estimating direction of arrivals of unknown deterministic wide-band signals incident from different directions onto a passive array. For the required signal estimation, alternative regularized least squares estimation techniques are proposed with significant improvement over the standard least squares techniques. Also, for the angle of arrival estimation of a large number of signals, a tree structured EM algorithm is proposed and compared with the conventional EM approach. Extensive simulation results are presented for comparison of the proposed algorithms with the current high-resolution methods of wide-band direction finding. In order to handle efficiently the case of available parametric prior models on the received waveforms, the required modifications are also given.

Keywords : Array Signal Processing, Source Localization, Wide-Band Direction Finding, Maximum Likelihood Estimation, Signal Parameter Estimation, EM Algorithm, Tree-Structured EM Algorithm.

ÖZET

GENİŞBANTLI İŞARETLERİN YAYILIM YÖNLERİNİN BEKLENTİ ENBÜYÜKLEME (BE) YÖNTEMİ İLE EN OLASI KESTİRİMİ

Nail Çadallı

Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans

Tez yöneticisi: Yard. Doç. Dr. Orhan Arıkan

Temmuz 1996

Alıcı bir dizi duyaca değişik yönlerden gelen genişbantlı kaynak işaretlerinin yayılım yönlerinin en olası kestirimi için Beklenti Enbüyükme (BE) yönteminin tam bir sunumu yapılmıştır. Gerekli işaret kestirimi konusunda enküçük kareler yönteminden daha iyi ve gürbüz sonuçlar verebilen düzenlenmiş enküçük kareler çözümleri önerilmiştir. Ayrıca, çok sayıda kaynak olması durumunda, yayılım yönlerinin bulunabilmesi için Dallı Beklenti Enbüyükme (DBE) yöntemi önerilmiş ve BE yöntemi ile karşılaştırılmıştır. Başarımlarını sınamak ve değerlendirmek üzere, bu yöntemler varolan genişbantlı yüksek çözünürlüklü yöntemlerle de karşılaştırılmıştır. Kaynak işaretleri hakkında önceden bilgi veya modellerin olması durumunda yöntemlerde yapılması gerekli değişiklikler de verilmiştir.

Anahtar Sözcükler : Dizisel İşaret İşleme, Kaynak Yer Kestirimi, Genişbantlı Yayılım Yönü Kestirimi, Enbüyük Olabilirlik Kestirimi (En Olası Kestirim), Sinyal Değişkenlerini Bulma, Beklenti Enbüyükme Yöntemi, Dallı Beklenti Enbüyükme Yöntemi.

ACKNOWLEDGMENTS

I would like to take this opportunity to record my deep gratitude to my supervisor Dr. Orhan Arıkan for his guidance, suggestions, and invaluable encouragement throughout the development of this thesis. I also wish to express that working with Dr. Orhan Arıkan has been a rewarding and a satisfying experience for me.

I am also deeply indebted to Dr. Billur Barshan and Dr. A. Enis Çetin for their reading and commenting on the thesis.

Many thanks to all my friends for their understanding.

Finally, I would like to thank to my parents and my sister for their patience, encouragement and constant support throughout my studies.

TABLE OF CONTENTS

1	INTRODUCTION	1
2	THE DATA MODEL	4
3	ML ESTIMATION VIA EM ALGORITHM	7
4	TREE-STRUCTURED EM ALGORITHM	19
5	SIMULATION RESULTS	22
5.1	Simulation Set I	22
5.2	Simulation Set II	25
5.3	Simulation Set III	26
6	CONCLUSIONS	36
	APPENDICES	44
A	Statistics of the Observations	44
B	Conditional Multivariate Normal Density	46
C	Form of the CRLB Expression	48

LIST OF FIGURES

3.1	Block diagram of the EM algorithm at the n 'th iteration.	18
4.1	Binary tree structure for the example case of 7 sources.	21
5.1	(a) Directional and (b) frequency dependent gain of a sensor. . . .	27
5.2	(a) DOA estimation error of the EM algorithm together with CRLB and result of TCT. (b) Magnified view of EM result with different signal estimation methods.	28
5.3	Signal estimation performance of the EM algorithm with different signal estimation methods.	29
5.4	Signal estimation error of the methods LS, RGLS and LS-SET. . .	30
5.5	True signals (dashed line) and best signal estimates (solid line) of the methods (a) LS, (b) RGLS-1, (c) RGLS-2, and (d) LS-SET among the realizations.	31
5.6	True signals (dashed line) and worst signal estimates (solid line) of the methods (a) LS, (b) RGLS-1, (c) RGLS-2, and (d) LS-SET among the realizations.	32
5.7	Performance of the EM algorithm with different number of sensors in the array. (a) Direction of arrival estimation error, (b) signal estimation error.	33
5.8	Trace of direction of arrival error.	34
5.9	Convergence comparison of Tree-Structured EM algorithm and conventional EM algorithm.	35

LIST OF TABLES

3.1	Steps of the EM algorithm.	16
3.2	Computational complexity of the signal estimation methods per iteration per frequency bin.	17

Chapter 1

INTRODUCTION

In many areas such as sonar, radar, radio-astronomy, seismology and imaging, reception data acquired by an array of sensors are processed to obtain information about the source locations and the characteristics of the emitted signals [1]. Since the measurement model of the direction of arrival (DOA) estimation is common to other applications, the results obtained in this thesis are expected to be widely used. In this study, we are concerned with the direction of arrival estimation of superimposed wide-band signals within the Maximum Likelihood (ML) estimation criterion by using the Expectation Maximization (EM) algorithm which is an iterative method for finding ML estimates [2]-[4].

The problem of direction of arrival estimation has been studied extensively both for the narrow-band and wide-band cases. As a result of these studies, numerous methods have been proposed for the solution of the problem. For the narrow-band case, the widely used ones are beamforming [5]-[8], linear prediction [9], Capon and Pisarenko methods [10] and others [11]-[16]. Although the maximum likelihood (ML) estimates are the most preferable [17], due to its higher computational cost, ML approach has not found much use in practice except for a few exact ML procedures such as Iterative Quadratic ML algorithm, Alternating Projection algorithm and Expectation Maximization algorithm [18, 19]. Therefore, for the ML criterion, suboptimal algorithms are used such as Multiple Signal Classification (MUSIC) algorithm, and its improved version for coherent signals, the Method of Direction Estimation (MODE) which are large sample approximations to ML and are based on the eigendecomposition of spatial correlation

matrix [20]-[26].

For the wide-band case, in order to avoid even more complicated ML estimation, estimates based on combination of narrow-band solutions at each frequency bin have been proposed. But the aggregation of independent information relevant to each frequency bin does not significantly reduce the variance of the estimates. In order to approximate the coherent aggregation, different algorithms have been proposed in the literature [27]. The commonly used strategy is to focus the information of each bin onto a single subspace and perform a narrow-band solution there [28]-[38].

However, the superposition property of the data acquisition system can be exploited by using the Expectation Maximization algorithm so as to reduce greatly the complexity of the ML estimation since EM algorithm finds maximum likelihood estimates iteratively without actually computing the likelihood function. The derivation of the EM algorithm for the direction finding problem in the narrow-band case is available and it is also applied to wide-band signals [3]. In this thesis, following a more complete derivation of wide-band EM algorithm, we will investigate various possibilities in the estimation of signals yielding superior results than the original approach.

In EM formalism, the observation, *incomplete data* is obtained via a many-to-one mapping from the *complete data* space that includes signals which we would obtain as the sensor outputs if we were able to observe the effect of each source separately. The EM algorithm iterates between estimating the likelihood of the complete data using the incomplete data and the current parameter estimates (E-step) and maximizing the estimated log-likelihood function to obtain the updated parameter estimates (M-step). Under mild regularity conditions, the iterations of the EM algorithm converges to a stationary point of the observed log-likelihood function, where at each iteration the likelihood of the estimated parameters is increased [39, 40].

Two extensions to the EM algorithm are proposed. The first one known as Cascade EM, (CEM), uses an intermediate complete data specification between the complete and incomplete data of the conventional EM method. It was reported that the CEM converges more rapidly and needs fewer computations per iteration when compared with the EM algorithm [41]. The second variant of the EM algorithm is the Space Alternating Generalized EM (SAGE) algorithm, in

which the parameters are updated sequentially by alternating between several hidden data spaces, unlike the EM algorithm where parameters are updated simultaneously [42, 43]. In SAGE, the complete data spaces are organized such that the maximization step of the EM algorithm is performed in less informative data spaces providing faster convergence.

In the present study, for the estimation of unknown signals arriving from different directions to a passive array, alternative regularized estimation schemes to the common least squares solution are investigated. For this purpose two different methods are used. The first one is an adaptive Tikhonov type regularized least-squares (RGLS) estimation method, and the second one is an averaged least-squares estimation (LS-SET) method over a set of angles in a neighborhood of the nominal angles. When this is performed for each direction separately, the size of the set containing the angles is smaller and the method is called LS-RSET, as an abbreviation for LS on a reduced set. It has been demonstrated that when regularized methods are used in the estimation of the received signals, the EM algorithm has better convergence behavior. Also, motivated with the ideas of CEM and SAGE, a tree structured hierarchy is used for the description of the relation between the complete data space and the observations. It is shown that the performance increases with respect to the non-tree structured EM algorithm even for moderate number of signals. In order to handle efficiently the case of available parametric prior models on the received waveforms, the required modifications are also given.

The thesis is organized as follows: In Chapter 2, the data model used in the sequel is presented. In Chapter 3, first the maximum likelihood estimation problem is presented, then the EM algorithm is described both in statistical terms and in formulation to the problem at hand. Tree-structured EM algorithm is suggested in Chapter 4. Then in Chapter 5, an extensive set of simulation examples is presented for the performance evaluation of the improvements suggested for the EM algorithm in comparison with the Cramer-Rao lower bound (CRLB) and the results of a recently proposed eigendecomposition-based subspace method for wideband signals. Finally we conclude with Chapter 6. There are appendices to the thesis for the sake of fluency in the text and the ease of reference.

Chapter 2

THE DATA MODEL

In this work, we will investigate the wide-band direction finding in the absence of near-field sources. Although, in principle, the same methodology can be used for arbitrarily located sources, the added complexity of the data model makes the presentation of the ideas more difficult. A discussion on the data model with point sources located in the near field of a sensor array is given in [44].

The impinging individual wavefields of the far field sources have negligible curvature effect, hence, can be assumed to have the same direction of arrivals at the sensors. The superposition of the individual wavefields forms the total wavefield which is spatially sampled by the sensor array. In addition to the spatial sampling, the data is obtained after a uniform time sampling of the output of each sensor. In the following, we will mainly be concerned with the case of non-parametric and unknown but deterministic source signals. When a prior parametric model of the source signals is available, this additional information can readily be utilized within the proposed framework with more reliable direction of arrival estimates.

The number of the arriving signals are assumed to be known, that is, the detection phase is assumed to be performed accurately with the use of standard algorithms developed for this purpose [45]-[49]. Note that the number of signals arriving on the array may be more than the number of emitters in the case of multi-path propagation.

Here we will first consider the narrow-band observation model to emphasize the required modifications for the wide-band case which is detailed later. For the

case of M narrow-band signal wavefronts from the directions $\theta_1, \theta_2 \dots \theta_M$ incident onto an array of P sensors, which have known directional gain characteristics, the output of the i 'th sensor, which is sampled at N sampling points with a sampling interval of T_s , can be written as

$$y_i(t) = \sum_{l=1}^M a_i(\theta_l) s_l(t - \tau_i(\theta_l)) + u_i(t), \quad 1 \leq i \leq P, \quad t = 0, T_s, \dots, (N-1)T_s, \quad (2.1)$$

where $s_l(t)$ is the l 'th signal emitted from the direction θ_l , $u_i(t)$ is the noise at the i 'th sensor which is zero mean, spatially and temporally white circularly symmetric complex Gaussian noise, $a_i(\theta)$ is the directional gain of the i 'th sensor, and $\tau_i(\theta_l)$ is the time delay (with respect to the phase center of the array) of the signal with the direction of arrival equal to θ_l . Since the source signals are assumed to be narrow-band, the sensor outputs can be closely approximated as follows

$$y_i(t) = \sum_{l=1}^M a_i(\theta_l) s_l(t) e^{-j\omega\tau_i(\theta_l)} + u_i(t), \quad 1 \leq i \leq P, \quad t = 0, T_s, \dots, (N-1)T_s, \quad (2.2)$$

which is based on the fact that for the narrow-band case, the signal amplitude does not change much during the period of time in which the signal passes across the array [4, 15]. In that case the time delay can be written as only a phase difference resulting in (2.2). However, for wide-band signals, there are both phase and amplitude variations across the array for which the validity of the narrow-band measurement model can not be justified.

In wide-band signal model, a quite general expression for the i 'th sensor output is

$$y_i(t) = \sum_{l=1}^M a_i(t, \theta_l) * s_l(t - \tau_i(\theta_l)) + u_i(t), \quad 1 \leq i \leq P, \quad t = 0, T_s, \dots, (N-1)T_s, \quad (2.3)$$

where in this case $a_i(t, \theta_l)$ represents the frequency dependent sensor gain. Since, the sensor gain is usually defined as a multiplicative operator in the Fourier Transform domain, the corresponding time domain expression involves the convolutional form given as in (2.3). The corresponding sensor output in the frequency domain can be closely approximated by using DFT, yielding

$$Y_i(k) = \sum_{l=1}^M A_i(k, \theta_l) e^{-j\frac{2\pi k\tau_i(\theta_l)}{FT_s}} S_l(k) + U_i(k), \quad 1 \leq i \leq P, \quad 0 \leq k < F, \quad (2.4)$$

where $Y_i(k)$, $A_i(k, \theta_i)$, $S_i(k)$, $U_i(k)$ are the F -point discrete Fourier transformations (DFT) of $y_i(t)$, $a_i(t, \theta_i)$, $s_i(t)$, $u_i(t)$, respectively. Since DFT is a unitary transformation, the transformed noise is still a spatially and temporally white complex Gaussian noise sequence. The size of the DFT is chosen sufficiently large in order to make the linear and circular convolutions almost equivalent.

In the array signal processing literature, the sensor gains are usually chosen omnidirectional and independent of frequency [3, 4]. But the model set forth above is more realistic in the sense that sensors usually have gain characteristics which vary as a function of both direction and frequency [50, 51].

The following definitions are used to simplify the representation:

$$\begin{aligned}
\Theta &= [\theta_1 \ \theta_2 \ \dots \ \theta_M]^T \\
\mathbf{b}(k, \theta) &= [A_1(k, \theta)e^{-j\frac{2\pi k}{F}\frac{r_1(\theta)}{T_s}} \ \dots \ A_P(k, \theta)e^{-j\frac{2\pi k}{F}\frac{r_P(\theta)}{T_s}}]^T \\
\mathbf{B}(k, \Theta) &= [\mathbf{b}(k, \theta_1) \ \dots \ \mathbf{b}(k, \theta_M)] \\
\mathbf{S}(k) &= [S_1(k) \ \dots \ S_M(k)]^T \\
\mathbf{Y}(k) &= [Y_1(k) \ \dots \ Y_P(k)]^T \\
\mathbf{U}(k) &= [U_1(k) \ \dots \ U_P(k)]^T \\
\mathcal{S} &= [\mathbf{S}^T(0) \ \dots \ \mathbf{S}^T(F-1)]^T \\
\mathcal{Y} &= [\mathbf{Y}^T(0) \ \dots \ \mathbf{Y}^T(F-1)]^T \\
\mathcal{U} &= [\mathbf{U}^T(0) \ \dots \ \mathbf{U}^T(F-1)]^T \\
\mathcal{B}(\Theta) &= \text{diag} \{ \mathbf{B}(0, \Theta) \ \dots \ \mathbf{B}(F-1, \Theta) \} , \tag{2.5}
\end{aligned}$$

Using these definitions, (2.4) can be written as

$$\mathcal{Y} = \mathcal{B}(\Theta)\mathcal{S} + \mathcal{U} , \tag{2.6}$$

or equivalently as

$$\mathbf{Y}(k) = \mathbf{B}(k, \Theta)\mathbf{S}(k) + \mathbf{U}(k) \quad 0 \leq k < F , \tag{2.7}$$

This final compact form of the measurement relation, which is the same as the signal model of the Cramer-Rao Lower Bound formula in [23], is used in our derivations.

Chapter 3

ML ESTIMATION VIA EM ALGORITHM

In this chapter, first we discuss the maximum likelihood estimation of the direction of arrivals and point out the difficulties involved in the maximization of the likelihood function. Then, we present the most straightforward set up for the EM methodology to obtain efficiently the ML estimates.

As shown in Appendix A, probability distribution of the observations as a function of the unknown parameters is

$$f_{\mathcal{Y}}(\mathcal{Y}; \Theta, \mathcal{S}) = (\pi\sigma^2)^{-FP} \exp \left\{ -\frac{1}{\sigma^2} \sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)]^\dagger [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)] \right\}. \quad (3.1)$$

This probability density function is called *the likelihood function* of the unknown parameters Θ and \mathbf{S} [15, 52]. The Maximum Likelihood estimates for Θ and \mathbf{S} maximize the likelihood function for a given set of observations. Since the likelihood function here is of an exponential form and since the logarithm is a monotonic function, maximizing the logarithm of the likelihood function is computationally easier than maximizing the likelihood function itself. After discarding constant terms, the logarithm of the likelihood function becomes

$$\mathcal{L}(\Theta, \mathcal{S}; \mathcal{Y}) = - \sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)]^\dagger [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)] . \quad (3.2)$$

This function is referred to as *the likelihood function* in the sequel. In array signal processing applications, usually there is more importance attached to reliable estimation of the direction of arrival parameters, Θ . Formally, maximum likelihood estimate for Θ is given as:

$$\hat{\Theta} = \arg \max_{\Theta} \left\{ \max_{\mathbf{S}} \left\{ - \sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)]^\dagger [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)] \right\} \right\} . \quad (3.3)$$

where the inner maximization is achieved by choosing $\mathbf{S}(k)$ as

$$\hat{\mathbf{S}}(k) = [\mathbf{B}^\dagger(k, \Theta)\mathbf{B}(k, \Theta)]^{-1} \mathbf{B}^\dagger(k, \Theta)\mathbf{Y}(k), \quad (3.4)$$

which is the least squares solution for the signals for a given Θ . Hence by using (3.3), we obtain the following formal expression for $\hat{\Theta}$:

$$\hat{\Theta} = \arg \max_{\Theta} \left\{ - \sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\hat{\mathbf{S}}(k)]^\dagger [\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\hat{\mathbf{S}}(k)] \right\} \quad (3.5)$$

Expanding the summand and discarding the terms that are not dependent on Θ , we obtain

$$\begin{aligned} \hat{\Theta} &= \arg \max_{\Theta} \left\{ \sum_{k=0}^{F-1} \mathbf{Y}^\dagger(k) \mathbf{B}(k, \Theta) \left(\mathbf{B}^\dagger(k, \Theta) \mathbf{B}(k, \Theta) \right)^{-1} \mathbf{B}^\dagger(k, \Theta) \mathbf{Y}(k) \right\} \\ &= \arg \max_{\Theta} \left\{ \sum_{k=0}^{F-1} \mathbf{Y}^\dagger(k) \mathbf{W}^\dagger(k, \Theta) \mathbf{W}(k, \Theta) \mathbf{Y}(k) \right\} \\ &= \arg \max_{\Theta} \left\{ \sum_{k=0}^{F-1} \text{tr} \left[\mathbf{W}(k, \Theta) \mathbf{Y}(k) \mathbf{Y}^\dagger(k) \mathbf{W}^\dagger(k, \Theta) \right] \right\} , \end{aligned} \quad (3.6)$$

where

$$\mathbf{W}(k, \Theta) = \mathbf{B}(k, \Theta) \left(\mathbf{B}^\dagger(k, \Theta) \mathbf{B}(k, \Theta) \right)^{-1} \mathbf{B}^\dagger(k, \Theta) . \quad (3.7)$$

Direct maximization of (3.6) by using numerical search methods is not only computationally demanding but also due to the usually complicated local maxima structure of the likelihood function, it is not guaranteed to converge to the global maxima. The Expectation Maximization (EM) method of obtaining the maximum likelihood estimates has been proposed to overcome the difficulty by an either parallel or sequential iterative search in much lower dimensional parameter spaces [2]. The EM method requires the identification of the so called

complete data space from which a many-to-one mapping is done to the *incomplete data space*, that is, the observation space. The underlying relation between the observed data \mathcal{Y} and the complete data \mathcal{X} is a many-to-one mapping of the form

$$\mathcal{H}(\mathcal{X}) = \mathcal{Y} . \quad (3.8)$$

For the sake of notational simplicity let $\Phi = [\Theta^T \ S^T]^T$. The observed data, \mathcal{Y} , is a realization of the random variable $\tilde{\mathcal{Y}}$ with statistics $f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi)$. Similarly the complete data is represented by \mathcal{X} and the corresponding random variable by $\tilde{\mathcal{X}}$ with statistics $f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi)$. Hence the statistical relation between the complete and incomplete data can be written as

$$\int_{\Omega} f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi) d\mathcal{X} = f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi), \quad \Omega = \{\mathcal{X} \mid \mathcal{H}(\mathcal{X}) = \mathcal{Y}\} . \quad (3.9)$$

Also, the following conditional density relation holds

$$f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi) = f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi) f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi), \quad \forall \mathcal{H}(\mathcal{X}) = \mathcal{Y} . \quad (3.10)$$

Taking the natural logarithm of both sides of (3.10) yields

$$\log f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi) = \log f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi) - \log f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi), \quad \forall \mathcal{H}(\mathcal{X}) = \mathcal{Y} . \quad (3.11)$$

Multiplying both sides with $f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi')$, where Φ' is a particular parameter value, and integrating over Ω , we obtain

$$\begin{aligned} \int_{\Omega} f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi') \log f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi) d\mathcal{X} &= \\ \int_{\Omega} f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi') \log f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi) d\mathcal{X} & \\ - \int_{\Omega} f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi') \log f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi) d\mathcal{X} &. \end{aligned} \quad (3.12)$$

Since

$$\int_{\Omega} f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi') d\mathcal{X} = 1 , \quad (3.13)$$

we can write

$$\log f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi) = \mathcal{E} \{ \log f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi) | \mathcal{Y}; \Phi' \} - \mathcal{E} \{ \log f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi) | \mathcal{Y}; \Phi' \} . \quad (3.14)$$

Now let

$$\mathcal{K}(\Phi, \Phi') = \mathcal{E} \{ \log f_{\tilde{\mathcal{X}}}(\mathcal{X}; \Phi) | \mathcal{Y}; \Phi' \} , \quad (3.15)$$

$$\mathcal{V}(\Phi, \Phi') = \mathcal{E} \{ \log f_{\tilde{\mathcal{X}}|\tilde{\mathcal{Y}}}(\mathcal{X}|\mathcal{Y}; \Phi) | \mathcal{Y}; \Phi' \} , \quad (3.16)$$

$$\mathcal{L}(\Phi) = f_{\tilde{\mathcal{Y}}}(\mathcal{Y}; \Phi) = \mathcal{K}(\Phi, \Phi') - \mathcal{V}(\Phi, \Phi') . \quad (3.17)$$

Note that, in fact, $\mathcal{L}(\Phi) = \mathcal{L}(\Theta, \mathcal{S}; \mathcal{Y})$, only the argument is omitted for the sake of simplicity. By using the discrimination function and the theorem on its non-negativity [53], we can write

$$\int_{\Omega} f_{\tilde{x}|\tilde{y}}(\mathcal{X}|\mathcal{Y}; \Phi') \log \frac{f_{\tilde{x}|\tilde{y}}(\mathcal{X}|\mathcal{Y}; \Phi')}{f_{\tilde{x}|\tilde{y}}(\mathcal{X}|\mathcal{Y}; \Phi)} d\mathcal{X} \geq 0 \quad (3.18)$$

and using the expansion property of the logarithm function, the following inequality is obtained:

$$\mathcal{V}(\Phi', \Phi') \geq \mathcal{V}(\Phi, \Phi') \quad . \quad (3.19)$$

Using this in (3.17), a relation can be found between $\mathcal{K}(\Phi, \Phi')$ and $\mathcal{L}(\Phi)$ independent of $\mathcal{V}(\Phi, \Phi')$. If we find the parameter Φ such that

$$\mathcal{K}(\Phi, \Phi') > \mathcal{K}(\Phi', \Phi') \quad (3.20)$$

then, from (3.17), for that particular value of Φ the following holds,

$$\mathcal{L}(\Phi) > \mathcal{L}(\Phi') \quad . \quad (3.21)$$

In other words, maximizing $\mathcal{K}(\Phi, \Phi')$ with respect to Φ yields a more likely estimate than Φ' . This method of improvement on the current value of Φ' , at time instant n , can be repeated in the following recursive algorithm:

$$\text{E-step} : \quad \text{Compute} \quad \mathcal{K}(\Phi, \hat{\Phi}^n) \quad (3.22)$$

$$\text{M-step} : \quad \hat{\Phi}^{n+1} = \arg \max_{\Phi} \mathcal{K}(\Phi, \hat{\Phi}^n) \quad . \quad (3.23)$$

In these recursive steps, it is guaranteed that the likelihood of the estimates monotonically improves [2]. In our application, the most commonly used complete data specification is $\mathbf{X}_l(k) = [X_{1l}(k) \cdots X_{Pl}(k)]^T$ which is the spectrum of the signal that would be observed at the sensors if we were able to observe the effect of l 'th source separately. The mean of the complete data $\mathbf{X}_l(k)$ is $\mathbf{b}(k, \theta_l)S_l(k)$ and its statistics are determined by the additive noise. Then the many-to-one mapping for all sources from the complete data space to the incomplete data space can be written as

$$\mathbf{Y}(k) = \sum_{l=1}^M \mathbf{X}_l(k), \quad 0 \leq k < F \quad . \quad (3.24)$$

The corresponding log-likelihood function of the complete data is

$$\mathcal{L}_c(\Theta, \mathcal{S}; \mathcal{X}) = - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l(k) - \mathbf{b}(k, \theta_l) S_l(k)\|^2, \quad (3.25)$$

where the complete data vector is defined as $\mathcal{X} = [\mathbf{X}^T(0) \dots \mathbf{X}^T(F-1)]^T$ and $\mathbf{X}(k) = [\mathbf{X}^T_1(k) \dots \mathbf{X}^T_M(k)]^T$. Here, the observed signal is decomposed to M constituents, hence each term of the summation on l can be maximized separately. Therefore, in the estimation of θ_l and $S_l(k)$, only $\mathbf{X}_l(k)$ is used along with the observations. At the n 'th iteration of the EM algorithm, expectation step conditionally estimates the likelihood of the complete data. Maximization step then finds the maximizer of the estimated likelihood. As shown in the Appendix B, the corresponding expectation step for the above complete data specification is:

$$\begin{aligned} \mathbf{X}_l^n(k) &= \mathcal{E}\{\mathbf{X}_l(k) | \Theta^n, \mathbf{S}^n(k), \mathbf{Y}(k)\} \\ &= \mathbf{b}(k, \theta_l^n) S_l^n(k) + \frac{1}{M} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta^n) \mathbf{S}^n(k)] \\ &\quad 0 \leq k < F. \end{aligned} \quad (3.26)$$

In maximization step, the complete data likelihood which is formed by using $\mathbf{X}_l^n(k)$ is maximized with respect to θ_l and $S_l(k)$. The θ_l is updated by numerically solving the following optimization problem:

$$\Theta^{n+1} = \arg \max_{\Theta} \left\{ \max_{\mathcal{S}} \left\{ - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l^n(k) - \mathbf{b}(k, \theta_l) S_l(k)\|^2 \right\} \right\}, \quad (3.27)$$

where, as in (3.3), there are two maximization problems inside one another which must be simultaneously solved. Since, for a given θ_l value, the solution of the l 'th term in the inner maximization is

$$\begin{aligned} \hat{S}_l(k) &= [\mathbf{b}^\dagger(k, \theta_l) \mathbf{b}(k, \theta_l)]^{-1} \mathbf{b}^\dagger(k, \theta_l) \mathbf{X}_l^n(k) \\ &= \frac{\mathbf{b}^\dagger(k, \theta_l) \mathbf{X}_l^n(k)}{\|\mathbf{b}(k, \theta_l)\|^2}, \end{aligned} \quad (3.28)$$

inserting this expression into (3.27) and solving for the outer maximization, the update for the direction of arrivals can be obtained formally as

$$\Theta^{n+1} = \arg \max_{\Theta} \left\{ - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l^n(k) - \mathbf{b}(k, \theta_l) \hat{S}_l(k)\|^2 \right\} \quad (3.29)$$

Since the required optimization can be performed in M individual optimizations of the nonnegative terms of the summation on l , significant computational saving is achieved compared to the required maximization in (3.3). Also, in this form, the required optimization can be performed in parallel where efficient one-dimensional optimization via line search methods can be used in each parallel process as shown in Figure 3.1 [54]. At the n 'th iteration of the EM algorithm the update formulas are as follows:

$$\text{E-step} : \mathbf{X}_l^n(k) = \mathbf{b}(k, \theta_l^n) S_l^n(k) + \frac{1}{M} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta^n) \mathbf{S}^n(k)] \quad (3.30)$$

$$\text{M-step} : \theta_l^{n+1} = \arg \max_{\theta} \sum_{k=0}^{F-1} \frac{\mathbf{b}^\dagger(k, \theta) \mathbf{X}_l^n(k) \mathbf{X}_l^{\dagger n}(k) \mathbf{b}(k, \theta)}{\|\mathbf{b}(k, \theta)\|^2} \quad (3.31)$$

$$S_l^{n+1}(k) = \frac{\mathbf{b}^\dagger(k, \theta_l^{n+1}) \mathbf{X}_l^n(k)}{\|\mathbf{b}(k, \theta_l^{n+1})\|^2} \quad (3.32)$$

In the above algorithm, corresponding to the joint maximization in (3.27), direction of arrival estimation part of the maximization step is (3.31), whereas, (3.32) is the signal estimation stage which uses the least squares solution given in (3.28) which is to be called LS-EM from now on. In the maximization step, the direction of arrival and signal estimation phases can be performed either one after the other for each l separately, as in the conventional EM algorithm, or signal estimation stage can be performed after the direction of arrival estimation is completed for all l . Actually, for the latter case, Θ^{n+1} is available after (3.31) and can be inserted into (2.7). Then $\mathbf{S}^{n+1}(k)$ can be solved for by using a number of alternatives such as the least squares (LS) solution which can be performed to estimate all M signal waveforms at once by using all currently updated direction estimates, giving

$$\hat{\mathbf{S}}(k) = [\mathbf{B}^\dagger(k, \Theta) \mathbf{B}(k, \Theta)]^{-1} \mathbf{B}^\dagger(k, \Theta) \mathbf{Y}(k) \quad (3.33)$$

This LS estimation has been proposed as a generalization of the EM algorithm to speed up convergence [4]. However, since the array manifold matrix $\mathbf{B}(k, \Theta)$ is used instead of the steering vector $\mathbf{b}(k, \theta)$, the required inversion in (3.33) may cause numerical instability problems during the iterations of the algorithm, especially for the case of sources with small separation. One way to avoid this is the use of more robust regularized least squares (RGLS) estimate:

$$\hat{\mathbf{S}}(k) = [\mathbf{B}^\dagger(k, \Theta) \mathbf{B}(k, \Theta) + \mu \mathbf{I}]^{-1} \mathbf{B}^\dagger(k, \Theta) \mathbf{Y}(k) \quad (3.34)$$

It is important to properly choose the regularization parameter μ . Some commonly used methods for this purpose were investigated [55, 56]. In our implementations, two adaptive strategies for the choice of μ are employed. First one, which is referred to as RGLS-1, works with the known noise variance, whereas the other one, RGLS-2, can be used when the noise variance is unknown [57].

An alternative to the adaptively regularized least squares estimates for source signals is the following which is referred to as LS-SET solution,

$$\begin{aligned}\hat{\mathbf{S}}(k) &= \arg \min_{\mathbf{S}(k)} \int_{\mathcal{Z}} \|\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)\|^2 d\Theta \\ &= \left[\int_{\mathcal{Z}} \mathbf{B}(k, \Theta)^\dagger \mathbf{B}(k, \Theta) d\Theta \right]^{-1} \left[\int_{\mathcal{Z}} \mathbf{B}(k, \Theta)^\dagger d\Theta \right] \mathbf{Y}(k),\end{aligned}\quad (3.35)$$

where \mathcal{Z} is a set of directions in a neighborhood of Θ . Since in the cost function of LS-SET, we use an average penalty in the neighborhood of the estimated direction of arrivals, the LS-SET solution provides signal estimates which is robust to the inaccuracies in the direction of arrival estimates. For the case of discrete neighborhood directions the integrals reduce to summations over an M -dimensional grid of L directions in each dimension. The number of required nested summations over an M dimensional grid in the above expression increases with the dimension of the direction vector Θ . Thus, the computational complexity increases exponentially with M , being in the order of L^M . Therefore, practically, this alternative, if used as above, is not preferable for large grid sizes and large number of superposed signals. However, since (3.35) is solely an alternative for the least squares solution, it can also be applied to the estimation in (3.28), that is, as in LS-EM, performing the estimation in each complete data space separately as follows:

$$\begin{aligned}\hat{S}_l(k) &= \arg \min_{S_l(k)} \int_{\mathcal{Z}_l} \|\mathbf{X}_l(k) - \mathbf{b}(k, \theta_l)S_l(k)\|^2 d\theta_l \\ &= \left[\int_{\mathcal{Z}_l} \mathbf{b}^\dagger(k, \theta_l) \mathbf{b}(k, \theta_l) d\theta_l \right]^{-1} \left[\int_{\mathcal{Z}_l} \mathbf{b}^\dagger(k, \theta_l) d\theta_l \right] \mathbf{X}_l(k).\end{aligned}\quad (3.36)$$

This solution is named as LS-RSET, as an abbreviation for LS on a reduced set, since the size of the set is much smaller. In that case, the summation terms are not nested inside each other, hence the computational load increases only linearly with M , being in the order of $L \times M$.

EM algorithm starts with $n = 0$ at which time Θ^0 is available, obtained by using a rough estimation. To find $\mathbf{X}_l^0(k)$ in (3.26), S_l^0 is needed and it is

estimated by using one of the methods of LS, RGLS or LS-SET. In that stage LS-EM and LS-RSET can not be used since they use the complete data in the estimation which is not available before the signal decomposition is performed. Then comes the expectation step where the observations are decomposed into complete data. In the maximization step, updates for the direction of arrivals and the signal values are obtained. For the signal estimation stage of this step, different alternatives other than LS-EM can be used such as LS, RGLS, LS-SET and LS-RSET. If it is desired to use the same signal estimation method throughout the process, the same method as in the initial signal estimation should be used in the signal estimation stage of the maximization step. For LS-RSET and LS-EM, initial signal estimation by using LS-SET or LS respectively may be preferable. The steps of the EM algorithm, explicitly showing the proper places to apply various alternatives, are summarized in Table 3.1.

The computational load (per iteration per frequency bin) of the alternative methods for signal estimation is shown in Table 3.2. J is the maximum number of iterations, usually in the order of 10, for the adaptive routine in the RGLS method which finds the optimum μ in (3.34). RGLS method may be computationally more intensive than LS for the case of comparable M and P values. However, in practice, the regularization parameter μ , which is determined for each frequency bin and given direction of arrival, can be computed a priori and stored in a lookup table, providing significant computational saving over the repeated use of the EM iterations thereafter. Furthermore, for large sized arrays and few sources, that is, for the case of large P and small M values, RGLS may even be computationally preferable without any need for a lookup table. For large number of sources, LS-SET can also be expensive in computation depending on the number of neighboring directions in the set. Then, LS-RSET is a proper substitute for such a case, since it uses lower-dimensional data. For the whole picture of the computational complexity per iteration of the EM algorithm, one should add the computational complexity of the expectation step, the direction of arrival update stage of the maximization step and the initial signal estimation step as well.

In conventional EM approach and the variants proposed in this chapter, the source signals are modeled to be arbitrary band-limited functions of time. It is appropriate to take this course in the absence of any a priori information for the

source signals. However, in many practical cases the source signals are known except for a few unknown parameters. For example, in the case of active direction finding applications, the received waveforms are scaled, phase and frequency shifted versions of the transmitted waveforms. Another example is the case of sources emitting signals which are known to be in a class of signals with a few members. For these cases, it is desirable to obtain maximum likelihood estimates for the direction of arrivals of the received signals as well as the unknown parameters of the signal itself. For this purpose, all of the previously given EM approaches can be used with the required modification in the maximization stage.

The output of the sensors can be written in terms of the parametric signal as

$$\mathbf{Y}(k) = \mathbf{B}(k, \Theta)\mathbf{S}(k, \Gamma) + \mathbf{U}(k) \quad 0 \leq k < F, \quad (3.37)$$

where $\Gamma = [\gamma_1 \dots \gamma_M]$ is the parameter set, each element of which is associated with the l 'th source signal. Each parameter γ may be a vector containing more than one element such as amplitude, phase, etc. of the signal. With the parametric signal model the maximization can be written as

$$\Theta^{n+1} = \arg \max_{\Theta} \left\{ \max_{\Gamma} \left\{ - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l^n(k) - \mathbf{b}(k, \theta_l) S_l(k, \gamma_l)\|^2 \right\} \right\} \quad (3.38)$$

For the inner maximization, a parametric optimization is employed to find a signal estimate $\hat{\mathbf{S}}(k, \Gamma)$ for a given value of Θ and this is used in the outer maximization as

$$\Theta^{n+1} = \arg \max_{\Theta} \left\{ - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l^n(k) - \mathbf{b}(k, \theta_l) \hat{S}_l(k, \gamma_l)\|^2 \right\} \quad (3.39)$$

Depending on the parametric optimization used to find $\hat{\mathbf{S}}(k, \Gamma)$, the maximization step may be easy or difficult to compute. With this required change, (Tree-Structured) EM algorithm is applicable to the parametric signal case. The estimation of the signal by using a parametric estimator should also be done efficiently, since the effect of the efficient signal estimation on the performance of the EM algorithm is demonstrated.

Initialization

1. Set $n = 0$.
2. Obtain a rough estimate for Θ^0 .
3. Using Θ^0 , solve (2.7) by LS, RGLS or LS-SET for $S^0(k)$

E-Step

4. Decompose the observation to complete data by using (3.26)

M-Step

5. Update Θ^n by using (3.31) for all l . If LS-EM or LS-RSET is to be used in Step 6, θ_l^n can be updated only.
6. Update $S^n(k)$ if LS, RGLS or LS-SET is used. Update $S_l^n(k)$ if LS-EM or LS-RSET is used. Do the latter for each l if Θ^n is updated in Step 5, otherwise goto Step 5 for the update of θ_{l+1}^n .

Loop

7. Increment n by 1 and continue from Step 4.

Table 3.1: Steps of the EM algorithm.

<i>Method</i>	<i>Multiplication/Division</i>	<i>Addition</i>
LS	$\frac{M^3}{3} + M^2(P + 1) + M(P - \frac{1}{3})$	$\frac{M^3}{3} + M^2(P + \frac{1}{2}) + M(P - \frac{13}{6})$
LS-EM	$2PM$	$2M(P - 1)$
RGLS-1	$MP + M^3 + 3M^2$ $+M(6 + 7J) + 5 + 2J$	$MP + M^3 + 2M^2$ $+4JM - 2J - 1$
RGLS-2	$M^3 + 2M^2 + (M + 1)P$ $M(9 + 7J) + 2J + 5$	$M^3 + 2M^2 + (M + 1)P$ $+M(4J + 1) - 2J - 1$
LS-SET	$(L^M - 1)M^2P + \frac{M^3}{3} + M^2$ $+M(P - \frac{1}{3})$	$(L^M - 1)M^2P + L^M MP + \frac{M^3}{3}$ $\frac{3}{2}M^2 - 2M$
LS-RSET	$M(L + 1)P + M$	$2MPL - MP - M$

Table 3.2: Computational complexity of the signal estimation methods per iteration per frequency bin.

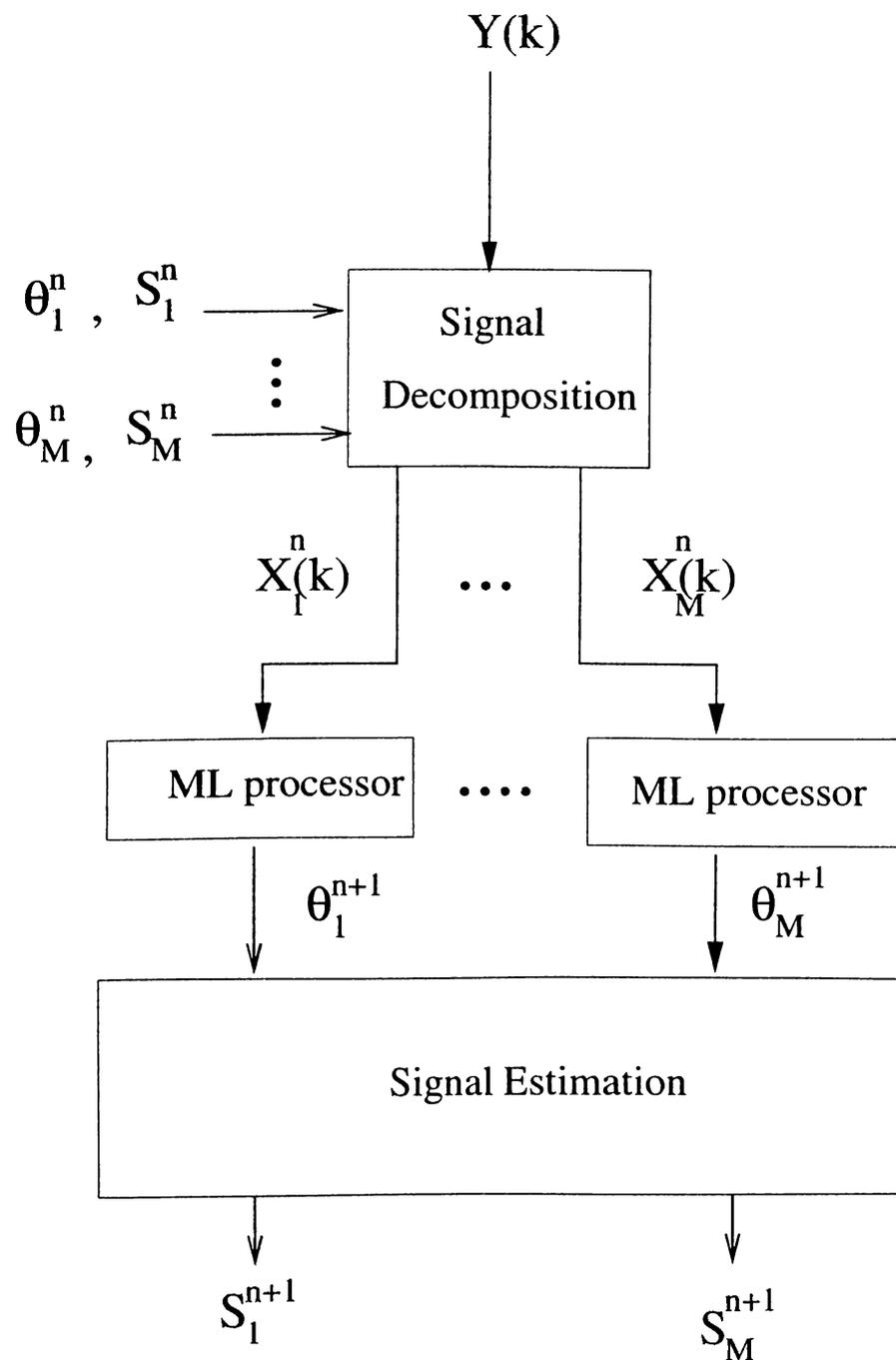


Figure 3.1: Block diagram of the EM algorithm at the n 'th iteration.

Chapter 4

TREE-STRUCTURED EM ALGORITHM

In this chapter, we propose to use a multi-level tree structured mapping between incomplete and complete data spaces rather than the commonly used data set up for the EM algorithm presented in Section 3. In this way we aim to bring together the superior features of two former extensions of the EM algorithm, namely, the cascade EM (CEM) algorithm, and the Space Alternating Generalized EM (SAGE) algorithm [41]-[43]. In CEM, an intermediate data specification between the complete and the incomplete data of the conventional EM method is used and intermediate EM steps at some iterations are performed. As a result, faster convergence with fewer computations per iteration is achieved when compared with the conventional EM algorithm. In SAGE, the parameters are sequentially updated by alternating between several hidden data spaces, unlike the EM algorithm where the parameters are updated simultaneously. The sequential maximization of the expected likelihood function in each hidden data space has been reported to be the main factor of the superior performance of SAGE compared to the conventional EM algorithm. With the purpose of capturing beneficial features of these algorithms, we propose to use a multilevel data hierarchy as shown in Figure 4.1 for the example case of 7 sources. The leaves of the tree hosts the complete data $\mathbf{X}_l(k)$'s and the root of the tree denotes the observation $\mathbf{Y}(k)$. The intermediate nodes correspond to the partial conditional incomplete data, $\mathbf{Y}_{i,j}(k)$, which are also updated during the iterations. It should

be noted that the intermediate data at a particular node is not simply obtained by summing the complete data of the leaves. The relevant branching indicates the relationship of the complete and incomplete data of the EM algorithm. The associated precise data assignment for the intermediate incomplete data is explained below.

In this setting, EM algorithm can be run for two sources at a time using the incomplete data at the joint node of two leaves which is obtained by using the intermediate data at the upper branch node and complete data which is not to be updated by the current run. For instance, to run EM algorithm for $\mathbf{X}_1(k)$ and $\mathbf{X}_2(k)$ we form the required incomplete data as

$$\mathbf{Y}_{1,1}(k) = \mathbf{Y}(k) - \sum_{l=5}^7 \mathbf{b}(k, \theta_l) S_l(k) , \quad (4.1)$$

$$\mathbf{Y}_{2,1}(k) = \mathbf{Y}_{1,1}(k) - \sum_{l=3}^4 \mathbf{b}(k, \theta_l) S_l(k) , \quad (4.2)$$

where $S_l(k)$ and θ_l are the current values of the signal and direction of arrival respectively. After a number of iterations are performed on the branch of $\mathbf{Y}_{2,1}(k)$, EM algorithm can be run for the branch of $\mathbf{Y}_{2,2}(k)$ which can be found as follows

$$\mathbf{Y}_{2,2}(k) = \mathbf{Y}_{1,1}(k) - \sum_{l=1}^2 \mathbf{b}(k, \theta_l) S_l(k) , \quad (4.3)$$

where, in this case, $S_l(k)$ and θ_l for $l = 1, 2$ are the updated values by the EM algorithm applied to the branch of $\mathbf{Y}_{2,1}(k)$ before. This switching may be repeated a number of times or until a convergence criterion is satisfied. Having updated values of signal and direction of arrivals for $l = 1 \dots 4$, the EM routine can be run for the branch of $\mathbf{Y}_{1,2}(k)$ with the same strategy of data assignment. The switching between the branches of $\mathbf{Y}_{1,1}(k)$ and $\mathbf{Y}_{1,2}(k)$ can be repeated too. In this tree structure, the branches can be seen as hidden data spaces that can be associated with the SAGE in the sense that not all of the parameters are updated at a time but only a subset of parameters are updated sequentially. Working in smaller dimensional spaces provides not only speed in convergence but also computational saving since in the tree structure, two sources are treated at a time with smaller dimensional matrices resulting in less number of computations per iteration. Therefore, it is suitable to use tree structure especially for large

number of sources and computationally expensive regularized signal estimation methods.

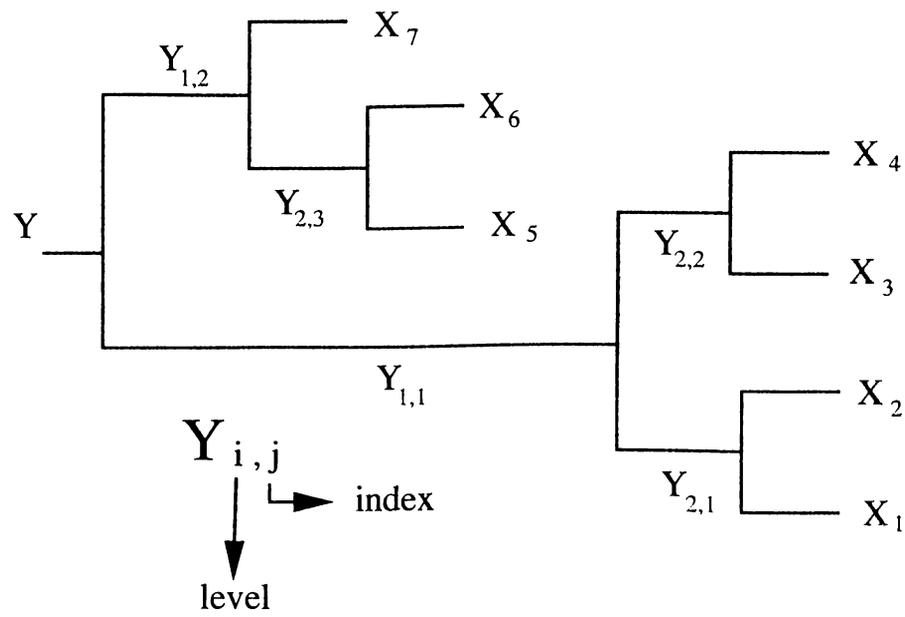


Figure 4.1: Binary tree structure for the example case of 7 sources.

Chapter 5

SIMULATION RESULTS

In this chapter, we compare the proposed algorithms both with each other and with the conventional EM approach. Also, we investigate the performance improvement with respect to one of the most improved subspace based wide-band direction finding algorithm known as Two Sided Correlation Transformation (TCT) [38].

5.1 Simulation Set I

This part of the simulations include the performance evaluation of the EM algorithm when, in the signal estimation stage of the maximization step, one of the proposed or the former signal estimation methods is used. The methods are termed as LS-EM, LS, RGLS, LS-SET and LS-RSET corresponding to (3.28), (3.33), (3.34), (3.35) and (3.36) respectively. The scenario is as follows: 2 wide-band signals with true direction of arrivals $\bar{\Theta} = [35^\circ \ -20^\circ]^T$ with respect to the normal of the array, are incident onto an array of 19 sensors. The number of the sensors is chosen by using the result of another simulation which is done to reveal the performance of the EM algorithm with changing sensor number. This simulation is detailed later. The array consists of colinear sensors which are coplanar with the arriving signals. This corresponds to the direction finding in a 2-D space where the direction of arrival parameters for each direction are simply scalars of the azimuthal angle. However, the formulation in the previous chapters are valid for the 3-D case in which there is also an elevation angle component in

the direction of arrival parameter θ_l , which is a vector rather than a scalar in that case, corresponding to the l 'th signal. A discussion on the required extension and the associated data model is given in [58]. The gain characteristics of the sensors are taken the same for all sensors which is also not a necessity. The directional and frequency dependence of the sensors are shown in Figure 5.1. Measurement noise in each sensor is assumed to be independent identically distributed circularly symmetric complex white Gaussian noise. Signals are taken as coherent linear FM waveforms with bandwidth larger than the center frequency.

For each of the signal estimation alternatives, the EM algorithm is run, within a convergence criterion and a bound on the maximum number of iterations, with an initial direction of arrival estimate $\Theta^0 = [32^\circ \ -17^\circ]^T$ for 10 realizations and *signal-to-noise ratio* ranging from -10 dB to 10 dB with 5 dB increments. Signal-to-noise ratio in decibels is defined as

$$SNR = 10 \log_{10} \frac{\text{energy of the noiseless observations}}{\text{variance of the noise}} \text{ dB} . \quad (5.1)$$

For initialization of the procedure, the initial signal estimation is performed by using RGLS-2 which is chosen by using the result of another simulation which is also detailed later. The performance is evaluated in terms of both the direction of arrival and the signal estimation error. The error for the direction of arrival estimation is

$$\epsilon_\Theta = \|\hat{\Theta} - \bar{\Theta}\|^2 , \quad (5.2)$$

where $\bar{\Theta}$ is the true direction of arrival and $\hat{\Theta}$ is its estimate. For the signal estimation, error is defined in terms of time domain signal values as

$$\epsilon_s = \frac{1}{MN} \sum_{l=1}^M \sum_{t=0}^{N-1} |\hat{s}_l(t) - \bar{s}_l(t)|^2 , \quad (5.3)$$

where $\hat{s}_l(t)$ is the estimated time domain signal and $\bar{s}_l(t)$ is the true signal waveform.

The direction of arrival estimation error of the EM algorithm using different signal estimation methods is shown in Figure 5.2-a together with the Cramer-Rao lower bound on the variance of the estimates and the results produced by the TCT algorithm. A magnified view of the results due to the signal estimation methods is shown in Figure 5.2-b. Shown in Figure 5.3 is the corresponding signal estimation error for each method.

As seen in Figure 5.2, the performance of the RGLS method, which is almost the same for RGLS-1 and RGLS-2 algorithms, seems to be superior to the others at all tried SNR values. This behavior is apparent especially at low SNR values, where the performance of LS degrades due to the numerical instability effect which is mentioned earlier. LS-SET and LS-RSET having almost identical behavior throughout the range of SNR values, also do better than LS or LS-EM for low SNR regions. Correspondingly, it is clear from Figure 5.3 that the RGLS method is significantly superior in producing signal estimates. The LS-SET performs slightly better than LS-RSET, both of which produce less error than the LS or LS-EM which are almost on the same curve. From these two figures, it can be said that, the signal and direction of arrival estimations are closely related. The TCT algorithm has not produced good results in the simulations. This may be due to the large bandwidth of the signals used. Also, the large number of frequency bins used in our simulations may cause a difficulty in focusing different signal subspaces associated with particular frequency bins to a common focusing subspace.

As mentioned earlier, a separate simulation is performed for the comparison of the methods LS, RGLS and LS-SET for the initial signal estimation as a solution to (2.7) with Θ^0 at hand. In order to adequately evaluate the signal estimation performance of the methods in case of initial direction of arrival error, the signal estimation is performed from the observations at SNR=0 dB, on a grid of directions corresponding to different deviations from the true values of each of the two direction of arrivals. The grid consists of 13 divisions for each dimension with equal deviation of 3° on each side of the true values. The error of signal estimation for each grid point corresponding to a pair of angles is plotted in Figure 5.4. The missing points in the surface of LS mesh, are due to the blow up of the method at those points. LS can show instable behavior in the case of ill-conditioned matrix even if the sources are not so close to each other. Actually, this is the reason that it needs regularization. On the other hand, regularized solution methods have estimated the signal robustly. RGLS-1 and RGLS-2 methods, which are close to each other in performance, produced better results than LS-SET as well. However, it is clear from the smoothness of the LS-SET mesh that, this method is more robust to inaccuracies in the initial direction of arrival. The best estimates of signals among the realizations are plotted in

Figure 5.5 and the worst estimates in Figure 5.6. This provides a comparison on the limits of the estimation performance of the corresponding methods. It is clear that, LS-SET can produce acceptable results even at the worst case where the performance of others degraded.

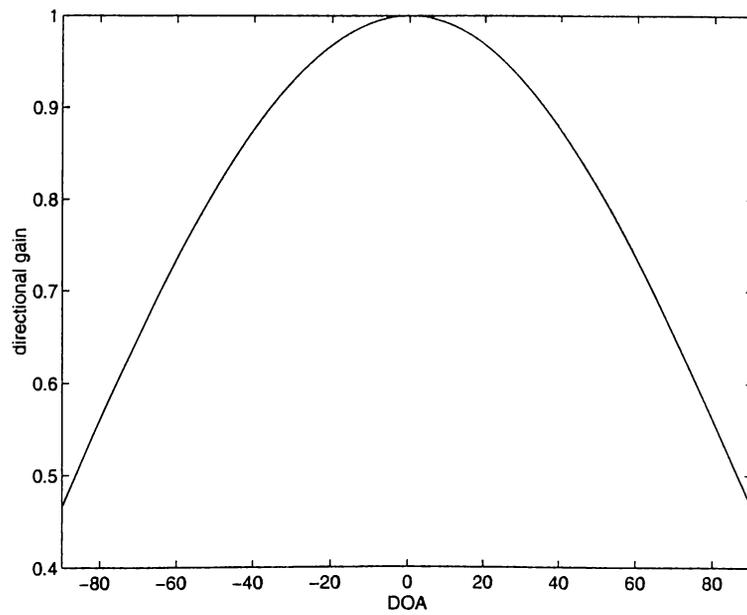
For determining the number of sensors to be used in the simulations, a separate simulation about the effect of the sensor number on the estimation is carried out. In that simulation the sensor number is increased from 3 to 31 by increments of 4 and the array is held symmetric around the origin. The direction of arrival and signal estimation error for the method RGLS-1 at an SNR value of 0 dB, as a function of the sensor number is shown in Figure 5.7. According to these results the sensor number is chosen to be 19 since there is insignificant decrease in the error after that point.

5.2 Simulation Set II

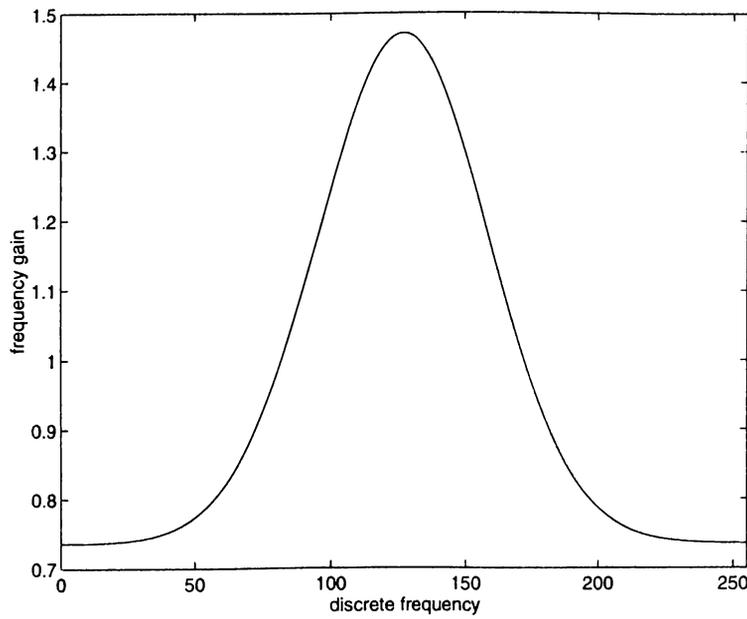
Here, as a complement to the previous section, the convergence behavior of the EM algorithm is to be evaluated which employs the signal estimation alternatives. The configuration is the same as in the previous section. The simulation is done for SNR=0 dB and the direction of arrival error is traced as the EM algorithm iterates. The result shown in Figure 5.8 is the average error of direction of arrival estimates over 10 realizations as a function of iteration number of the EM algorithm. The convergence of RGLS methods clearly outperforms the others, providing a significant gain in the number of iterations which should be performed to reach a satisfactory convergence level. The convergence behaviors should also be taken into account together with the performance of the methods outlined in the previous set of simulations and the complexity of the algorithm tabulated in Chapter 3, in order to decide which method to use in a particular application where the criterion of the selection may be computational complexity, robustness, speed of convergence or all of them.

5.3 Simulation Set III

In this part of the simulations, the tree-structured EM (TSEM) algorithm is compared with the conventional EM algorithm for the case of 4 sources from directions $\Theta = [35^\circ \ -20^\circ \ -50^\circ \ 60^\circ]^T$ at SNR=0dB. Initial directions are given as $\Theta_0 = [32^\circ \ -47^\circ \ -17^\circ \ 47^\circ]^T$. In TSEM, the EM algorithm, which uses the LS-RSET in its signal estimation, is run for two sources at a time, with a maximum number of iterations of 5 at each branch. After 5 iterations at one branch, the algorithm switches to the other branch. This switching is also repeated 8 times. Therefore, in total, 40 iterations of EM algorithm are performed for each direction of arrival. For comparison purposes, the EM algorithm is also run for 4 sources with a maximum number of iterations of 40. The direction of arrival estimation error is traced throughout the processes and displayed in Figure 5.9 where it can be clearly seen that the Tree-Structured EM algorithm has a significantly high speed of convergence. Also, note that beyond this speed, tree structure provides saving in computation in each iteration due to the smaller size of matrices and vectors, handled in the algorithm.

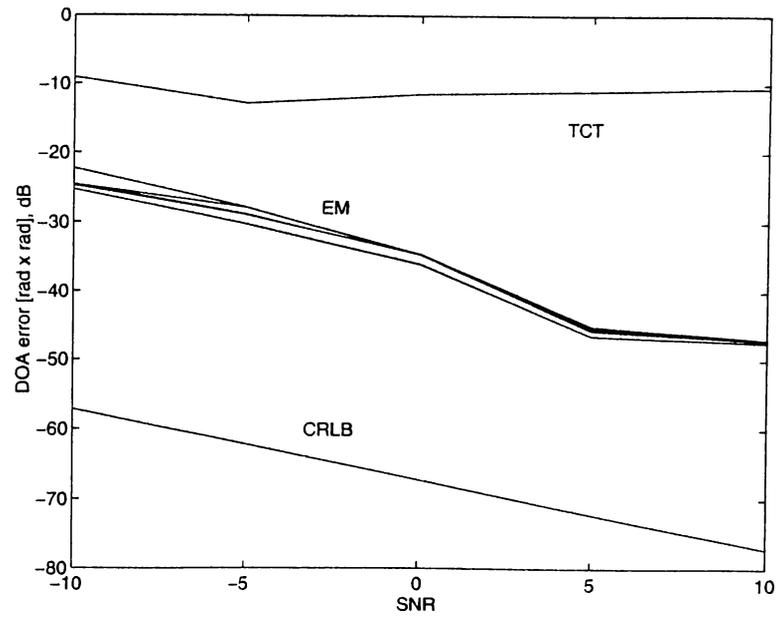


(a)

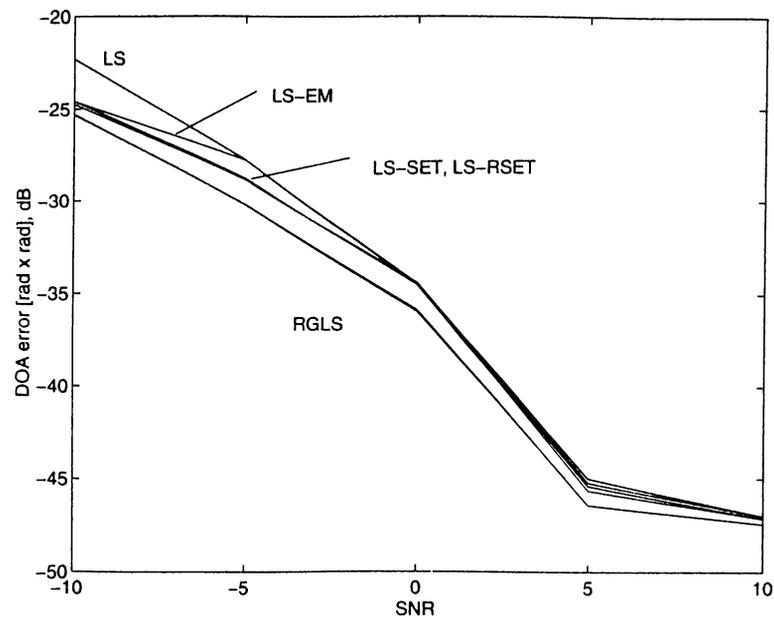


(b)

Figure 5.1: (a) Directional and (b) frequency dependent gain of a sensor.



(a)



(b)

Figure 5.2: (a) DOA estimation error of the EM algorithm together with CRLB and result of TCT. (b) Magnified view of EM result with different signal estimation methods.

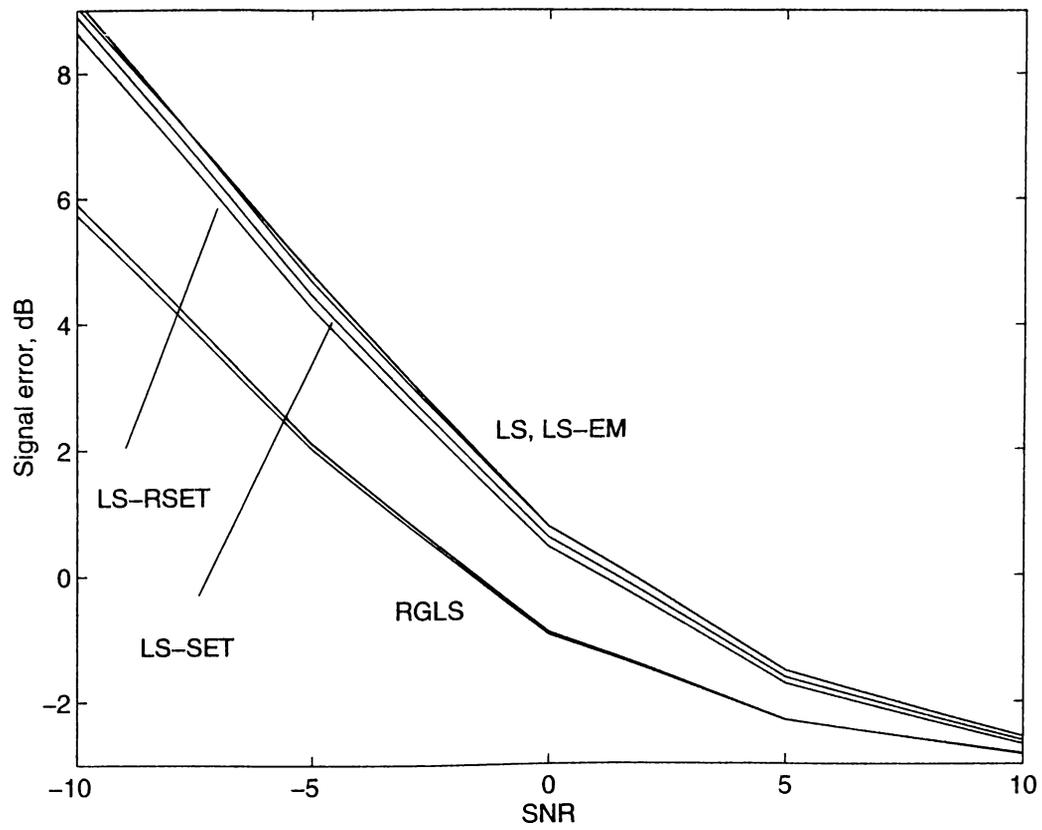


Figure 5.3: Signal estimation performance of the EM algorithm with different signal estimation methods.

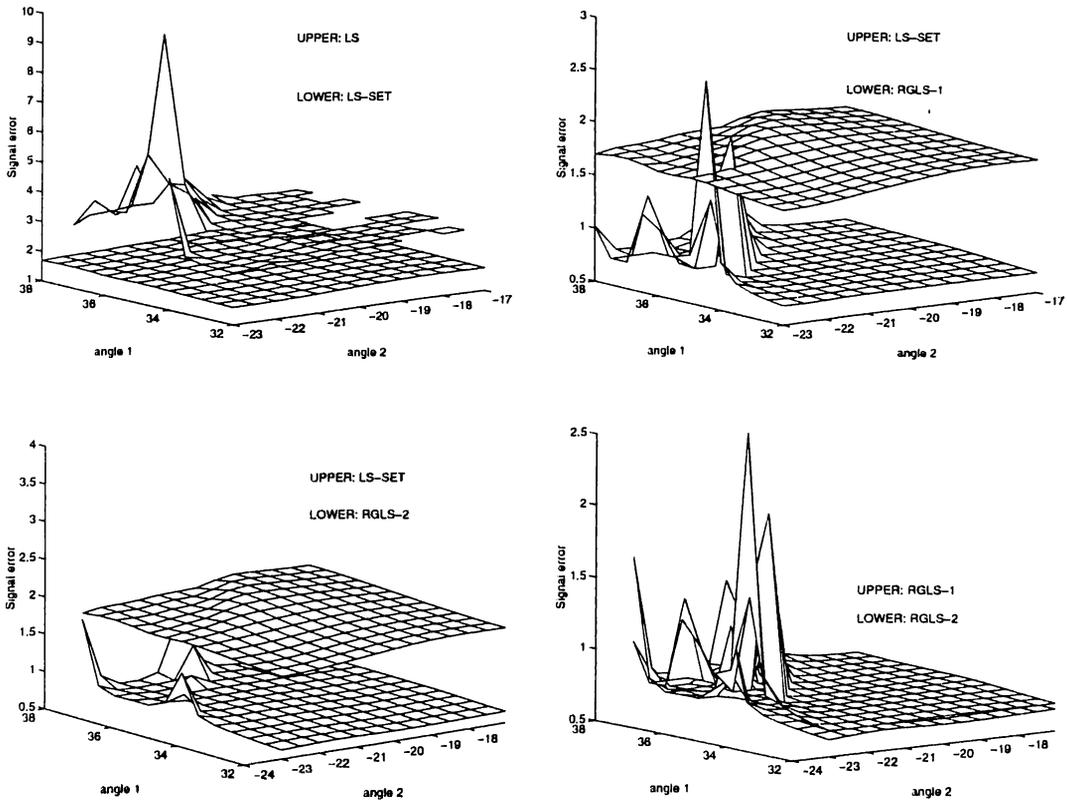
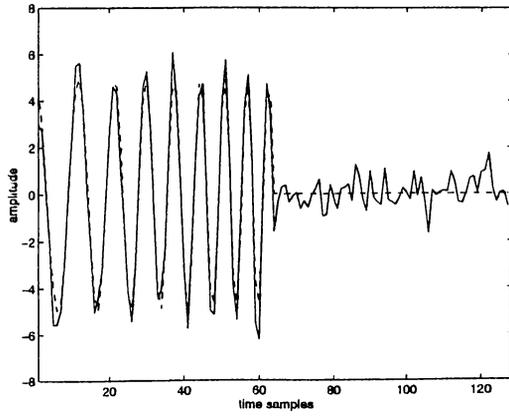
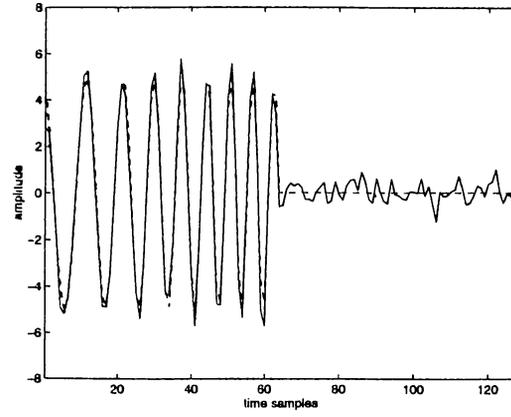


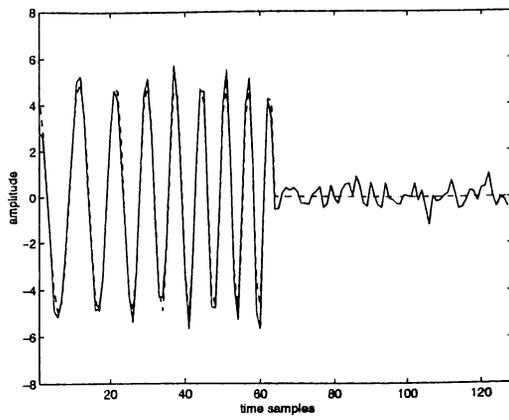
Figure 5.4: Signal estimation error of the methods LS, RGLS and LS-SET.



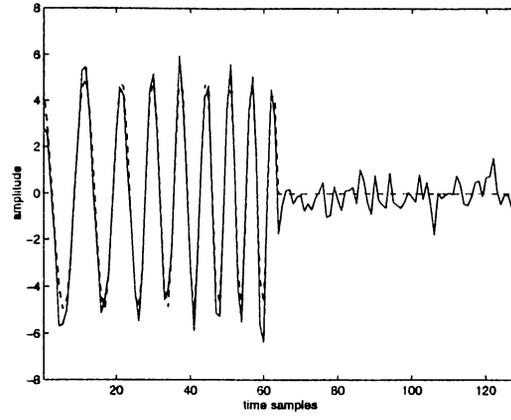
(a)



(b)

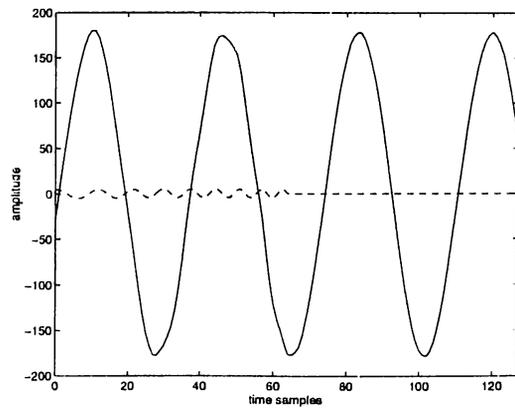


(c)

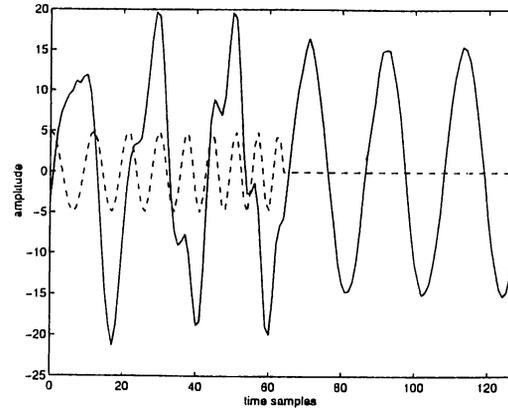


(d)

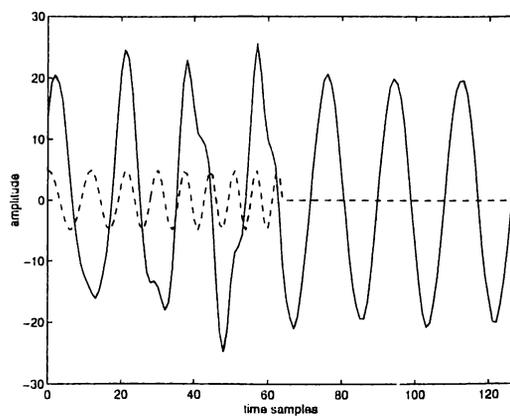
Figure 5.5: True signals (dashed line) and best signal estimates (solid line) of the methods (a) LS, (b) RGLS-1, (c) RGLS-2, and (d) LS-SET among the realizations.



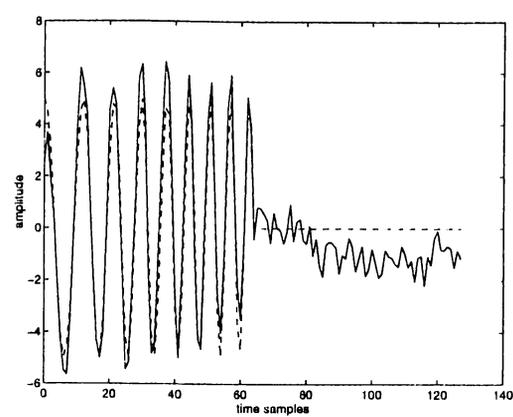
(a)



(b)

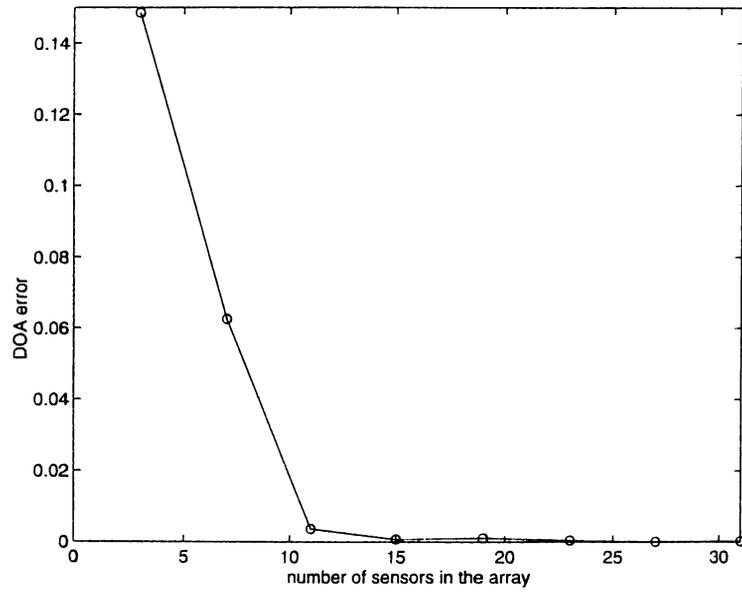


(c)

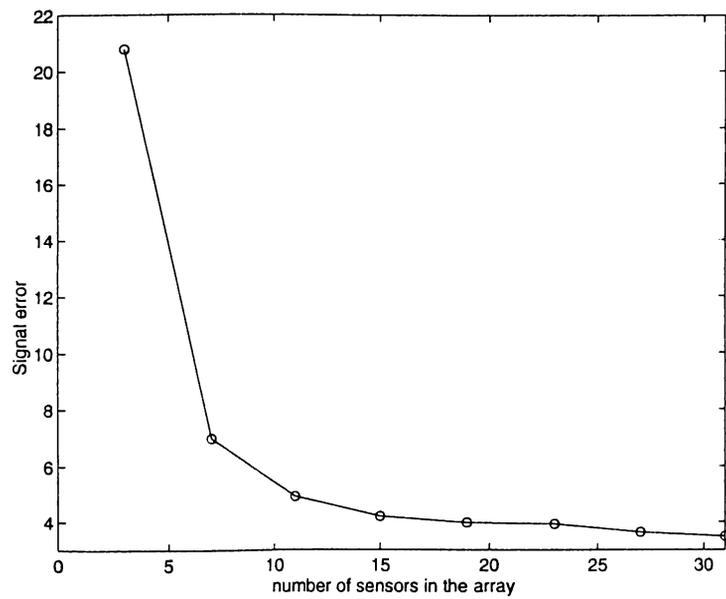


(d)

Figure 5.6: True signals (dashed line) and worst signal estimates (solid line) of the methods (a) LS, (b) RGLS-1, (c) RGLS-2, and (d) LS-SET among the realizations.



(a)



(b)

Figure 5.7: Performance of the EM algorithm with different number of sensors in the array. (a) Direction of arrival estimation error, (b) signal estimation error.

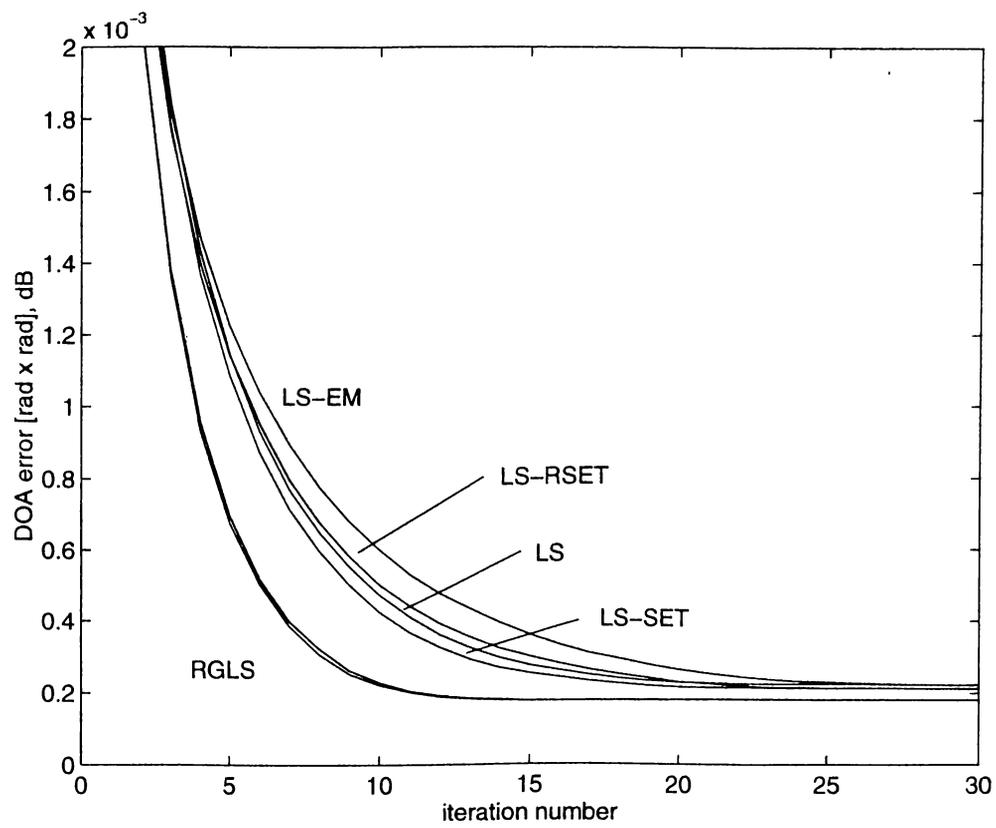


Figure 5.8: Trace of direction of arrival error.

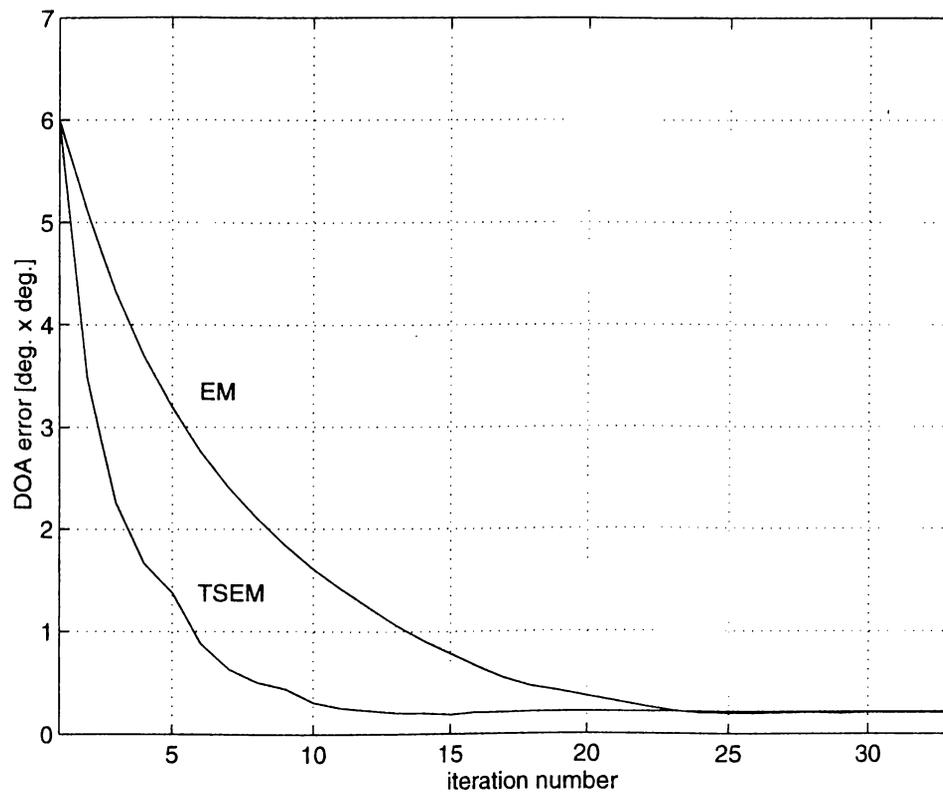


Figure 5.9: Convergence comparison of Tree-Structured EM algorithm and conventional EM algorithm.

Chapter 6

CONCLUSIONS

A thorough derivation of the Expectation Maximization algorithm, which is an iterative numerical method of Maximum Likelihood estimation, is presented for the case of estimating direction of arrivals and waveforms of unknown deterministic wide-band signals incident from different directions onto a passive array of sensors. Also, the required modifications in the algorithm is given for the case of an available prior parametric model on the received signal waveforms.

To improve the accuracy, alternative regularized least squares estimation techniques are proposed to replace the common least squares solution for estimation of incident signal waveforms. By the comparison of the proposed signal estimation alternatives both with each other and with the common least squares solution, it is shown that better estimates for the signal waveforms can be obtained by using regularized signal estimation methods.

In order to obtain reliable estimates for the direction of arrivals even at low signal-to-noise ratio and speed up the convergence of the EM algorithm, proposed alternative regularized signal estimation methods are applied to both the initial signal estimation step and the signal estimation stage of the maximization step in the EM algorithm. Also, it is shown that with better signal estimation during the iterations, EM algorithm converges faster to more reliable direction of arrival and signal estimates compared with the conventional EM algorithm.

As a generalization of the commonly used direct mapping, a binary tree structured multiple level data mapping is proposed between the observations and the

complete data of the EM algorithm. It is demonstrated that the proposed Tree-Structured EM (TSEM) algorithm converges faster than the conventional EM algorithm does even for moderate number of sources. In the binary tree structure, the EM algorithm can be run for two sources at a time with lower-dimensional data. Hence, in addition to speeding up the convergence, the tree structure provides considerable saving in computation, especially for large number of superposed signals.

REFERENCES

- [1] S. Haykin, ed. *Array Signal Processing*. Prentice-Hall, 1985.
- [2] A. P. Dempster, N. M. Laird, and D. B. Rubin “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Stat. Soc.*, vol. B-39, pp. 1–37, 1977.
- [3] M. Feder and E. Weinstein “Parameter estimation of superimposed signals using the EM algorithm,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 477–489, Apr. 1988.
- [4] M. I. Miller and D. R. Fuhrmann “Maximum-likelihood narrow-band direction finding and the EM algorithm,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1560–1577, Sep. 1990.
- [5] D. E. Dudgeon “Fundamentals of digital array processing,” *Proc. IEEE*, vol. 65, pp. 898–904, June 1977.
- [6] H. Fan, E. I. El-Masry, and W.K. Jenkins “Resolution enhancement of digital beamformers,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1041–1052, Oct. 1984.
- [7] R.A. Mucci “A comparison of efficient beamforming algorithms,” *IEEE Trans. Acoust., Speech, Signal Processing*. vol. ASSP-32, pp. 548–558, June 1984.
- [8] B. Widrow and S. D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, 1985.
- [9] J. Makhoul “Linear prediction: A tutorial review,” *Proc. IEEE*, vol. 63, pp. 561–680, 1975.

- [10] D. H. Johnson “The application of spectral estimation methods to bearing estimation problems.” *Proc. IEEE*, vol. 70, pp. 1018–1028, Sep. 1982.
- [11] J. Munier and G. Y. Delisle “Spatial analysis in passive listening using adaptive techniques,” *Proc. IEEE*, vol. 75, pp. 1458–1471, Nov. 1987.
- [12] U. Nickel “Angular superresolution with phased array radar : a review of algorithms and operational constraints,” *Proc. IEE*, vol. 134, Pt. F, pp. 53–59, Feb. 1987.
- [13] H. Wang and M. Kaveh “On the performance of signal subspace processing Part I: Narrow-band systems,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1201–1209, Oct. 1986.
- [14] S. U. Pillai. *Array Signal Processing*. Springer-Verlag, 1989.
- [15] S. Haykin, J. Litva, and T.J. Shepherd, eds. *Radar Array Processing*. Springer series in information sciences. Springer-Verlag, 1993.
- [16] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, 1993.
- [17] M. Wax and T. Kailath “Optimum localization of multiple sources by passive arrays,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1210–1218, Oct. 1983.
- [18] Y. Bresler and A. Macovski “Exact maximum likelihood parameter estimation of superimposed exponential signals in noise,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1081–1089, Oct. 1986.
- [19] I. Ziskind and M. Wax “Maximum likelihood localization of multiple sources by alternating projection,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.
- [20] G. Bienvenu and L. Kopp “Optimality of high resolution array processing using the eigensystem approach,” *IEEE Trans. Signal Processing*, vol. ASSP-31, pp. 1235–1248, Oct. 1983.

- [21] R. O. Schmidt "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.
- [22] D. R. Farrier, D.J. Jeffries, and R. Mardani "Theoretical performance prediction of the MUSIC algorithm," *Proc. IEE*, vol. 135, Pt. F, June 1988.
- [23] P. Stoica and A. Nehorai "MUSIC, Maximum Likelihood and Cramer-Rao Bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
- [24] P. Stoica and A. Nehorai "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.
- [25] P. Stoica and A. Nehorai "MUSIC, Maximum Likelihood and Cramer-Rao Bound : Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2140–2150, Dec. 1990.
- [26] P. Stoica and C. Sharman "Maximum likelihood methods for direction of arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1132–1143, July 1990.
- [27] Y. Grenier "Wideband source location through frequency dependent modeling," *IEEE Trans. Signal Processing*, vol. 42, pp. 1087–1096, May 1994.
- [28] G. Su and M. Morf "The signal subspace approach for multiple wide-band emitter location," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1502–1522, Dec. 1983.
- [29] M. Wax, T. Shan, and T. Kailath "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 817–827, Aug. 1984.
- [30] H. Wang and M. Kaveh "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 823–831, Aug. 1985.

- [31] K. M. Buckley and L. J. Griffiths “Broad-band signal subspace spatial spectrum (BASS-ALE) estimation,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 953–964, July 1988.
- [32] H. Hung and M. Kaveh “Focussing matrices for coherent signal subspace processing,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1272–1281, Aug. 1988.
- [33] J. Krolik and D. Swingler “Multiple broad-band source location using steered covariance matrices,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1481–1494, Oct. 1989.
- [34] H. Hung and M. Kaveh “Coherent wide-band ESPRIT method for direction of arrival estimation of multiple wide-band sources,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 354–360, Feb. 1990.
- [35] M. A. Doron and A. J. Weiss “On focusing matrices for wide-band array processing,” *IEEE Trans. Signal Processing*, vol. 40, pp. 1295–1302, June 1992.
- [36] M. Allam and A. Moghaddamjoo “Spatial-temporal DFT projection for wideband array processing,” *IEEE Signal Process. Lett.*, vol. 1, pp. 35–37, Feb. 1994.
- [37] T. Lee “Efficient wideband source localization using beamforming invariance technique,” *IEEE Trans. Signal Processing*, vol. 42, pp. 1376–1387, June 1994.
- [38] S. Valaee and P. Kabal “Wideband array processing using a two-sided correlation transformation,” *IEEE Trans. Signal Processing*, vol. 43, pp. 160–172, Jan. 1995.
- [39] R. A. Boyles “On the convergence of the EM algorithm,” *J. Roy. Stat. Soc.*, vol. B-45, pp. 47–50, 1983.
- [40] C. F. J. Wu “On the convergence properties of the EM algorithm,” *The Annals of Statistics*, vol. 11, pp. 95–103, 1983.

- [41] M. Segal and E. Weinstein “The Cascade EM algorithm,” *Proc. IEEE*, vol. 76, pp. 1388–1390, Oct. 1988.
- [42] J. A. Fessler and A. O. Hero “Space alternating generalized expectation-maximization algorithm,” *IEEE Trans. Signal Processing*, vol. 42, pp. 2664–2677, Oct. 1994.
- [43] J. A. Fessler and A. O. Hero “Penalized maximum-likelihood image reconstruction using space alternating generalized EM algorithms,” *IEEE Trans. Image Processing*, vol. 4, pp. 1417–1429, Oct. 1995.
- [44] J. Lee, Y. Chen, and C. Yeh “A covariance approximation method for near-field direction finding using a uniform linear array,” *IEEE Trans. Signal Processing*, vol. 43, pp. 1293–1298, May 1995.
- [45] Q. Wu and D. R. Fuhrmann “A parametric method for determining the number of signals,” in *Proc. 26th Allerton Conf., Commun., Contr., Computing*, pp. 219–228. University of Illinois, Sep. 1988.
- [46] Y. Wu and K. W. Tam “Randomized penalty in detection of the number of signals,” *IEEE Trans. Signal Processing*, vol. 42, pp. 2692–2696, Oct. 1994.
- [47] J. Rissanen “Modeling by shortest data description,” *Automatica*, vol. 14, pp. 465–471, 1978.
- [48] M. Wax and T. Kailath “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, Apr. 1985.
- [49] D. B. Williams “Counting the degrees of freedom when using AIC and MDL to detect signals,” *IEEE Trans. Signal Processing*, vol. 42, pp. 3282–3284, Nov. 1994.
- [50] E. A. Wolff. *Antenna Analysis*. Artech House, Norwood, MA, 1988.
- [51] R. J. Mailloux. *Phased Array Antenna Handbook*. Artech House, 1994.
- [52] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall signal processing series. Prentice-Hall, 1993.

- [53] R. E. Blahut. *Principles and practice of information theory*. Addison Wesley, 1987.
- [54] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 2nd edition, 1984.
- [55] A. Hoerl and W. Kennard “Ridge regression: biased estimation for non-orthogonal problems,” *Technometrics*, vol. 12, pp. 55–68, 1970.
- [56] H. D. Vinod “Survey of ridge regression and related techniques for improvements over ordinary least squares,” *Review of Economics and Statistics*, vol. 60, pp. 121–131, 1978.
- [57] P. A. V. B. Swamy, J. S. Mehta, and P. N. Rappoport “Two methods of evaluating Hoerl and Kennard’s ridge regression,” *Commun. Statist. - Theor. Meth.*, vol. A7, pp. 1133–1155, 1978.
- [58] F. A. Sakarya and M. H. Hayes “Estimating 2-D DOA angles using nonlinear array configurations,” *IEEE Trans. Signal Processing*, vol. 43, pp. 2212–2216, Sep. 1995.
- [59] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.
- [60] N. R. Goodman “Statistical analysis based on a certain multivariate complex distribution,” *Annals of Math. Statistics*, vol. 34, 1963.

Appendix A

Statistics of the Observations

Probability density function of a real Gaussian zero mean vector of size N (N 'th order real multivariate Gaussian density) can be written as the joint distribution of the real components of the random vector as a result of the definition of the joint normality [59] as follows,

$$f(\mathbf{r}) = f(r_1, r_2, \dots, r_N) = (2\pi)^{-\frac{N}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{r}^T \mathbf{C}^{-1} \mathbf{r} \right\} \quad (\text{A.1})$$

where $\mathbf{C} = \mathcal{E}\{\mathbf{r}\mathbf{r}^T\}$ is the covariance matrix. For a complex zero mean Gaussian vector $\mathbf{z} = \mathbf{u} + j\mathbf{v}$ of size N , its statistical properties are specified in terms of the joint density of its components,

$$\begin{aligned} f(\mathbf{z}) &= f(u_1, u_2, \dots, u_N, v_1, v_2, \dots, v_N) \\ &= (2\pi)^{-N} |\mathbf{D}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix} \mathbf{D}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \right\}, \end{aligned} \quad (\text{A.2})$$

where $2N^2 + N$ real parameters of correlations are held by the covariance matrix,

$$\mathbf{D} = \begin{bmatrix} \mathbf{C}_{\mathbf{u}\mathbf{u}} & \mathbf{C}_{\mathbf{u}\mathbf{v}} \\ \mathbf{C}_{\mathbf{v}\mathbf{u}} & \mathbf{C}_{\mathbf{v}\mathbf{v}} \end{bmatrix} \quad (\text{A.3})$$

On the other hand, the covariance of the complex vector \mathbf{z}

$$\mathbf{C}_{\mathbf{z}\mathbf{z}} = \mathbf{C}_{\mathbf{u}\mathbf{u}} + \mathbf{C}_{\mathbf{v}\mathbf{v}} - j(\mathbf{C}_{\mathbf{u}\mathbf{v}} - \mathbf{C}_{\mathbf{v}\mathbf{u}}), \quad (\text{A.4})$$

has N^2 real parameters. Thus, unlike the real case, the density of \mathbf{z} can not in general be determined in terms of $\mathbf{C}_{\mathbf{z}\mathbf{z}}$. A special class of complex normal vectors

that are statistically determined in terms of their covariance matrix is given by Goodman [60].

Theorem 1 ([59, 60]) *For the random vector $\mathbf{z} = \mathbf{u} + j\mathbf{v}$. if the vectors \mathbf{u} and \mathbf{v} are such that*

$$\mathbf{C}_{\mathbf{u}\mathbf{u}} = \mathbf{C}_{\mathbf{v}\mathbf{v}} \text{ and } \mathbf{C}_{\mathbf{u}\mathbf{v}} = -\mathbf{C}_{\mathbf{v}\mathbf{u}} , \quad (\text{A.5})$$

then

$$\mathbf{C}_{\mathbf{z}\mathbf{z}} = 2(\mathbf{C}_{\mathbf{u}\mathbf{u}} - j\mathbf{C}_{\mathbf{u}\mathbf{v}}) , \quad (\text{A.6})$$

and the density is given as

$$f(\mathbf{z}) = (\pi)^{-N} |\mathbf{C}_{\mathbf{z}\mathbf{z}}|^{-\frac{1}{2}} \exp \left\{ -\mathbf{z}^\dagger \mathbf{C}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{z} \right\} . \quad (\text{A.7})$$

□

If $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ then $\mathbf{u}, \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2} \mathbf{I})$. This is a special case of the theorem and the density can be written as follows,

$$f(\mathbf{z}) = f(\mathbf{u}, \mathbf{v}) = (2\pi)^{-N} \left(\frac{\sigma^2}{2}\right)^{-N} \exp \left\{ -\frac{1}{2} \mathbf{z}^\dagger \frac{2}{\sigma^2} \mathbf{z} \right\} = (\pi\sigma^2)^{-N} \exp \left\{ -\frac{1}{\sigma^2} \mathbf{z}^\dagger \mathbf{z} \right\} . \quad (\text{A.8})$$

In (2.7), $\mathbf{U}(k) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{P \times P})$. Therefore, the probability density function of the observations, where $\tilde{\mathbf{Y}}$ is the underlying random variable. is as follows,

$$f_{\tilde{\mathbf{Y}}}(\mathcal{Y} | \Theta, \mathcal{S}) = (\pi\sigma^2)^{-FP} \exp \left\{ -\frac{1}{\sigma^2} \sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta) \mathbf{S}(k)]^\dagger [\mathbf{Y}(k) - \mathbf{B}(k, \Theta) \mathbf{S}(k)] \right\} . \quad (\text{A.9})$$

Appendix B

Conditional Multivariate Normal Density

In this appendix, the conditional density function of a multivariate normal distribution is presented. Let \mathbf{x}_1 and \mathbf{x}_2 be jointly normal random vectors of distributions $\mathcal{N}(\mathbf{m}_1, \mathbf{C}_{11})$, $\mathcal{N}(\mathbf{m}_2, \mathbf{C}_{22})$ and orders N_1 and N_2 respectively. For real random vectors, the order is the length of the vector while for the complex case it is twice the length of the vector. The joint characteristics are expressed in terms of the vector $[\mathbf{x}_1^T \mathbf{x}_2^T]^T$, whose covariance matrix is

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}. \quad (\text{B.1})$$

The conditional distribution can be obtained by

$$f(\mathbf{x}_1|\mathbf{x}_2) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_2)}. \quad (\text{B.2})$$

The distribution of \mathbf{x}_2 is

$$f(\mathbf{x}_2) = (2\pi)^{-\frac{N_2}{2}} |\mathbf{C}_{22}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_2 - \mathbf{m}_2)^T \mathbf{C}_{22}^{-1} (\mathbf{x}_2 - \mathbf{m}_2) \right\}. \quad (\text{B.3})$$

The joint density $f(\mathbf{x}_1, \mathbf{x}_2)$ is to be expressed in terms of the submatrices of the matrix \mathbf{C} . From the inverse of a partitioned matrix,

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}_0^{-1} & -\mathbf{C}_0^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \\ -\mathbf{C}_{22}^{-1} \mathbf{C}_{12}^T \mathbf{C}_0^{-1} & \mathbf{C}_{22}^{-1} + \mathbf{C}_{22}^{-1} \mathbf{C}_{12}^T \mathbf{C}_0^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \end{bmatrix}, \quad (\text{B.4})$$

and

$$|\mathbf{C}| = |\mathbf{C}_{22}| \cdot |\mathbf{C}_0| , \quad (\text{B.5})$$

where

$$\mathbf{C}_0 = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{12}^T . \quad (\text{B.6})$$

Then the joint density can be written as

$$\begin{aligned} f(\mathbf{x}_1, \mathbf{x}_2) = & (2\pi)^{-\frac{N_1+N_2}{2}} |\mathbf{C}_{22}|^{-\frac{1}{2}} \cdot |\mathbf{C}_0|^{-\frac{1}{2}} \\ & \cdot \exp\left\{-\frac{1}{2}\left[(\mathbf{x}_1 - \mathbf{m}_1)^T \mathbf{C}_0^{-1}(\mathbf{x}_1 - \mathbf{m}_1) \right. \right. \\ & - (\mathbf{x}_2 - \mathbf{m}_2)^T \mathbf{C}_{22}^{-1} \mathbf{C}_{12}^T \mathbf{C}_0^{-1}(\mathbf{x}_1 - \mathbf{m}_1) \\ & - (\mathbf{x}_1 - \mathbf{m}_1)^T \mathbf{C}_0^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2) \\ & \left. \left. + (\mathbf{x}_2 - \mathbf{m}_2)^T \mathbf{C}_{22}^{-1} \mathbf{C}_{12}^T \mathbf{C}_0^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2) \right. \right. \\ & \left. \left. + (\mathbf{x}_2 - \mathbf{m}_2)^T \mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2)\right]\right\} . \quad (\text{B.7}) \end{aligned}$$

Upon division by $f(\mathbf{x}_2)$, the term $(2\pi)^{-\frac{N_2}{2}} |\mathbf{C}_{22}|^{-\frac{1}{2}}$ cancels and the last term of the exponent in (B.7) vanishes. The remaining terms can be written as

$$\begin{aligned} f(\mathbf{x}_1|\mathbf{x}_2) = & (2\pi)^{-\frac{N_1}{2}} |\mathbf{C}_0|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\mathbf{x}_1 - \mathbf{m}_1 - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2)\right]^T \right. \\ & \left. \cdot \mathbf{C}_0^{-1}\left[\mathbf{x}_1 - \mathbf{m}_1 - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2)\right]\right\} . \quad (\text{B.8}) \end{aligned}$$

This is the density of a multinormal distribution with mean vector

$$\mathbf{m}_1 + \mathbf{C}_{12}\mathbf{C}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2) , \quad (\text{B.9})$$

and covariance matrix \mathbf{C}_0 .

Appendix C

Form of the CRLB Expression

The Cramer-Rao Lower Bound for the unbiased estimates of the direction of arrivals for the narrow-band case has been derived [24, 25]. Here the wide-band data model is applied to this derivation and the theorem given in [24] is modified accordingly as the following:

Theorem 2

- A1 :* $P > M$ and $\mathbf{b}(k, \theta) \forall \theta$ are linearly independent.
A2 : $\mathcal{E}\{\mathbf{U}(k)\} = 0$
 $\mathcal{E}\{\mathbf{U}(k)\mathbf{U}^\dagger(k)\} = \sigma^2\mathbf{I}$
 $\mathcal{E}\{\mathbf{U}(k)\mathbf{U}^T(k)\} = 0$
AML : $\mathcal{E}\{\mathbf{U}(k)\mathbf{U}^\dagger(l)\} = \mathcal{E}\{\mathbf{U}(k)\mathbf{U}^T(l)\} = 0$ for $k \neq l$
and $\mathbf{U}(k)$ is Gaussian.

With the above assumptions, the Cramer-Rao Lower Bound for the parameter Θ is given as,

$$CRB(\Theta) = \frac{\sigma^2}{2} \left[\sum_{k=0}^{F-1} \mathcal{R}e \left\{ \mathbf{G}_k^\dagger \mathbf{D}^\dagger \left[\mathbf{I} - \mathbf{B}_k \left(\mathbf{B}_k^\dagger \mathbf{B}_k \right)^{-1} \right] \mathbf{B}_k^\dagger \right\} \mathbf{D} \mathbf{G}_k \right]^{-1} \quad (\text{C.1})$$

$$\text{var}_{CR}(\sigma^2) = \frac{\sigma^2}{PF}, \quad (\text{C.2})$$

where

$$\mathbf{B}_k = \mathbf{B}(k, \Theta) \quad (\text{C.3})$$

$$\mathbf{G}_k = \text{diag}\{S_1(k) \dots S_M(k)\} \quad (\text{C.4})$$

$$\mathbf{D} = [\mathbf{d}(\theta_1) \dots \mathbf{d}(\theta_M)]^T \quad (\text{C.5})$$

$$\mathbf{d}(\theta) = [d_1(\theta) \dots d_P(\theta)]^T = \frac{\partial \mathbf{b}(k, \theta)}{\partial \theta} = \left[\frac{\partial b_1(k, \theta)}{\partial \theta} \dots \frac{\partial b_P(k, \theta)}{\partial \theta} \right]^T. \quad (\text{C.6})$$

□

The derivative terms in our application can be evaluated as

$$d_i(\theta) = \frac{\partial b_i(k, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left\{ A_i(k, \theta) e^{-j \frac{2\pi k}{F} \frac{\tau_i(\theta)}{T_s}} \right\}. \quad (\text{C.7})$$

Since we take the dependence of the sensor gain separable in terms of frequency and direction, we can write

$$A_i(k, \theta) = A_i^f(k) A_i^d(\theta). \quad (\text{C.8})$$

Then,

$$d_i(\theta) = A_i^f(k) e^{-j \frac{2\pi k}{F} \frac{\tau_i(\theta)}{T_s}} \frac{\partial}{\partial \theta} A_i^d(\theta) + A_i(k, \theta) \left(-j \frac{2\pi k}{F} \frac{\tau_i(\theta)}{T_s} e^{-j \frac{2\pi k}{F} \frac{\tau_i(\theta)}{T_s}} \frac{\partial \tau_i(\theta)}{\partial \theta} \right). \quad (\text{C.9})$$