A Maximum Likelihood Narrowband Direction of Arrival Estimator

Supratim Saha*
Dept. of Electrical and Computer Engineering
University of Rhode Island
Steven Kay
Dept. of Electrical and Computer Engineering
University of Rhode Island
Kingston, RI 02881
401-874-2874(voice)
401-782-6422(fax)
sup@ele.uri.edu
1-DETC

April 17, 2001

keywords:DOA estimation, Monte Carlo methods, optimization methods,

Abstract

We present a new implementation for maximum likelihood (ML) estimation of direction of arrival (DOA) of multiple narrowband plane waves in noise. The proposed estimator can be applied to the estimation of DOAs of both deterministic and stochastic signals. In this paper we discuss the estimation of DOAs for stochastic signals only.

The proposed method uses a global optimization procedure to maximize the compressed likelihood function, which is a function of only the DOAs. We use importance sampling to obtain the global optimum DOAs. It is shown via simulations that the method allows for resolution of DOAs at low angular separation and high noise levels for which other suboptimal techniques fail. Furthermore the method does not suffer from guaranteed lack of convergence to a global maximum as do most of the reported iterative ML methods. We compare our simulation results to that of EM algorithm which is yet another implementation of the ML method of narrowband DOA estimation.

*Corresponding Author: Supratim Saha (email:sup@ele.uri.edu)

1 Introduction

Many problems in radar, sonar, seismology, etc. utilize the signals recorded via an array of sensors to perform high resolution estimation of the DOAs of emitters and the source amplitudes or powers. This problem has been studied extensively in recent years [1], [3]. The maximum likelihood (ML) technique was one of the first to be investigated. Nonetheless, because of the high computational load of the multivariate nonlinear maximization, it did not become popular. Its exact and most straightforward implementation requires a multidimensional grid search, whose computational complexity increases exponentially with the number of sources. Instead, the suboptimal techniques which make use of the properties of eigenvectors of the sample covariance matrix, have been dominant in the field owing to their simplicity of implementation. For a review of these techniques the reader is referred to [9], [10], [13].

The performance of these suboptimal eigenvector based techniques is inferior to that of the ML technique especially under important scenarios such as low signal to noise ratios or small angular separations or a few number of snapshots. Moreover, these techniques cannot handle the case of coherent signals, which appear in specular multipath propagation. The ML technique can perform well even when there is only one snapshot [8], whereas these techniques fail completely. As a result, in spite of the difficulty of implementation of the ML technique, research has not stopped for an alternative efficient implementation. Some of the implementations of the ML techniques reported for deterministic signals are [1], [4], [5], [8], and for the stochastic signal model [5], [6]. The methods proposed in [4], [5] are iterative. In [6], only a separable solution for the DOAs and the source covariance matrix has been proposed, with DOAs being determined with an impractical exhaustive grid search.

In this paper we present a *non-iterative* implementation of the ML estimator for the stochastic signal model. However, for either signal model, deterministic or stochastic, the joint ML estimation of the source DOAs and the complex amplitudes for the deterministic signal model, and the source covariance matrix for the stochastic signal model decouples. For the deterministic case, with a known narrowband signal waveform, the compressed likelihood function has the same form as in the case of sinusoidal parameter estimation [15]. Under the stochastic signal model, the likelihood function depends on the DOAs and all the elements of the arbitrary source covariance matrix. The compressed likelihood function for the stochastic signal model was derived by Jaffer [6] for an arbitrary source covariance matrix.

An ideal implementation of the global maximizer of the compressed likelihood function in either case requires a multidimensional grid search. We replace this step by using the theorem for global optimization discussed in [21]. It requires a multidimensional integration, which at first appears impractical. But such types of integrals can be well approximated by Monte Carlo techniques [22], [23]. In particular, importance sampling has been shown to be a very powerful Monte Carlo technique, allowing multidimensional integrals to be evaluated efficiently. We will use the importance sampling approach to obtain the DOA estimates.

The paper is organized as follows. In Section 2 we discuss the ML problem formulation for the problem. In Section 3, we deal with the compressed likelihood function which needs to be maximized in order to obtain the DOA estimates. In Section 4, the details of the proposed DOA estimator based on global optimization theorem [21] is discussed. In Section 5 and 6 the use of importance sampling to efficiently evaluate the multidimensional integral and obtain the DOAs estimate is described. Section 7 presents the summary of the algorithm. Section 8 contains some simulation results. Finally, in Section 9 we give conclusions.

2 Problem Formulation

Consider M narrowband signal sources located at angles $\theta_1, \ldots, \theta_M$, whose emitted signals are incident on a line array of P sensors indexed by $i = 0, \ldots, P-1$, as shown in figure 1. The possible range of the DOA angles is $0 \leq \theta_i \leq \pi, i = 1, \ldots, M$. The distances of the sensors are measured with respect to the leftmost sensor, which lies at the origin. We first consider an arbitrarily spaced array, where the distance of the i^{th} sensor from the origin is d_i, $i = 1, 2, \ldots, P-1$. The complex gain response of the i^{th} sensor to the k^{th} narrowband signal is $a_i(\theta_k)$. As discussed in the introduction, we consider signal waveforms which are known, except for complex amplitude. Thus, the k^{th} narrowband signal for the snapshot at instant t is given by $A_k(t)s_k(t)$, where $A_k(t)$ corresponds to the complex amplitude, which could be deterministic or random and $s_k(t)$ is the known narrowband signal. By the narrowband assumption, under the deterministic signal case, $A_k(t)$ is independent of t, and thus becomes a parameter which does not vary from snapshot to snapshot. However for the stochastic signal case, $A_k(t)$ does vary from snapshot to snapshot, but is assumed to be a realization of a random variable having the same probability density function (PDF) for all snapshots. In this paper $s_k(t)$ is of the form

$$s_k(t) = \exp(j2\pi F_k t)$$

where F_k is the center frequency of the k^{th} source, which is independent of time. However, $s_k(t)$ need not be in this form. The only requirement is that it should be fully known and narrowband, which could be satisfied by many other signals whose bandwidth is small compared to the known center frequency. As a result of the narrowband assumption the signal amplitude does not change appreciably during the period of time required to transit the array. Thus, the noisy signal received at the i^{th} sensor at the instant t is a superposition of the responses to the M individual signals, and is given by,

$$y_i(t) = \sum_{k=1}^{M} A_k(t) s_k\left(t - \tau_i(\theta_k)\right) a_i(\theta_k) + n_i(t) \tag{1}$$

where $\tau_i(\theta_k)$ refers to the delay experienced by the k^{th} signal in reaching the i^{th} sensor. From Fig. 1, $\tau_i(\theta_k) = -\frac{d_i \cos(\theta_k)}{c}$. Thus,

$$y_i(t) = \sum_{k=1}^{M} A_k(t) s_k \left(t + \frac{d_i \cos(\theta_k)}{c}\right) a_i(\theta_k) + n_i(t). \tag{2}$$

Since the signals are narrowband with center frequency of the k^{th} source being F_k, (2) can be written as

$$y_i(t) = \sum_{k=1}^{M} A_k(t) s_k(t) \exp\left(j2\pi \frac{d_i F_k \cos(\theta_k)}{c}\right) a_i(\theta_k) + n_i(t). \tag{3}$$

The above equation can be made more compact by defining the array vector $\mathbf{y}(t)$ for the t^{th} snapshot as

$$\mathbf{y}(t) = [y_0(t) \ \ldots \ y_{P-1}(t)]^T . \tag{4}$$

Further,

$$\boldsymbol{\theta} = [\theta_1 \ldots \theta_M]^T . \tag{5}$$

and

$$\mathbf{s}(t) = [A_1(t)s_1(t) \ldots A_M(t)s_M(t)]^T . \tag{6}$$

Further we assume that all the sources have the same narrowband center frequency F_c. Thus, $s_1(t) = \ldots = s_M(t) = s(t) = \exp(j2\pi F_c t)$, so that

$$\mathbf{s}(t) = [A_1(t) \ldots A_M(t)]^T s(t) = \mathbf{a}(t)s(t) \tag{7}$$

where $[A_1(t) \ldots A_M(t)]^T = \mathbf{a}(t)$. Let,

$$\mathbf{D}(\boldsymbol{\theta}) = [\mathbf{d}(\theta_1) \ldots \mathbf{d}(\theta_M)] \tag{8}$$

where $\mathbf{d}(\theta_k)$ is given by,

$$\mathbf{d}(\theta_k) = \left[\exp\left(j2\pi \frac{d_0 F_c \cos(\theta_k)}{c}\right) a_0(\theta_k) \ldots \exp\left(j2\pi \frac{d_{P-1} F_c \cos(\theta_k)}{c}\right) a_{P-1}(\theta_k)\right]^T . \tag{9}$$

It is assumed that the distances of the sensors from origin are known. Now if the sensors are uniformly spaced at a distance of $\frac{\lambda}{2}$ from each other, with $d_0 = 0$, and all the sensors have identical unity gain response for all the sources then $\mathbf{d}(\theta_k)$ reduces to

$$\mathbf{d}(\theta_k) = [1 \quad \exp(j\pi(1)\cos(\theta_k)) \quad \ldots \exp(j\pi(P-1)\cos(\theta_k))]^T . \tag{10}$$

Thus the basic equation (3), from the above assumptions, can be expressed in matrix form as

$$\mathbf{y}(t) = \mathbf{D}(\boldsymbol{\theta})\mathbf{a}(t)s(t) + \mathbf{n}(t) \tag{11}$$

where $\mathbf{y}(t)$ is the $P \times 1$ array vector for the t^{th} snapshot, $\mathbf{D}(\boldsymbol{\theta})$ is a $P \times M$ matrix defined in (8), $s(t)$ has been defined earlier and $\mathbf{n}(t)$ is the $P \times 1$ array noise vector. This is our basic data model. From the available vector $\mathbf{y}(t)$, at time instants $t = t_1, \ldots t_N$, often referred to as snapshots, the task is to determine the DOAs $\theta_1, \ldots \theta_M$. The processing strategies to determine the vector of angles $\boldsymbol{\theta}$ depend on further statistical properties of the signals and the noise. Throughout we will assume that the noise is *spatially* and *temporally white* [16].

For the signals themselves, there can be two different models. In one the signals are assumed to be known except for complex amplitude [4]. Note that this is not the same as in [1], [5], [8], where the vector $\mathbf{s}(t)$ is considered fully unknown for all snapshots. In the other signal model, the signal envelopes have a complex Gaussian PDF, with unknown covariance matrix, with successive samples being independent of each other. The M-dimensional amplitude vector $\mathbf{a}$ has a complex Gaussian distribution with zero mean and covariance matrix $\mathbf{P} = E[\mathbf{a}\mathbf{a}^H]$, which is independent of t. Note that the model equation is the same under either model. However the probability density functions for the observations are different for the two models, and hence they have different likelihood functions. We deal with only the stochastic signal model in this paper. However the method can be easily extended for deterministic signal case. For details of the implementation of the proposed method for deterministic signal case, the reader is referred to [20].

3 Compressed Likelihood Function

In the stochastic signal case, there are two different approaches used to derive the likelihood function [20]. In the first approach, it is assumed that the signal sources have a joint multidimensional Gaussian distribution with an unknown covariance matrix $\mathbf{P}$. No further assumptions are made on the structure of $\mathbf{P}$ [6]. In the other approach it is assumed that the source has a known prior Gaussian PDF with all the parameters of the PDF completely known. If this is the case, the likelihood function is obtained by integrating the product of the conditional PDF with respect to complex amplitudes and the prior PDF of the complex amplitudes. The integrated function thus becomes a function of only the unknown DOA vector $\boldsymbol{\theta}$. However, since it is more difficult to assign a prior PDF whose parameters are completely known, we adopt the former approach, with unknown covariance matrix $\mathbf{P}$.

Under the random Gaussian signal assumption, from (11), $\mathbf{y}(t_k)$ is a sum of independent Gaussian random vectors, $\mathbf{D}(\boldsymbol{\theta})\mathbf{s}(t_k)$ (with $\mathbf{s}(t)$ defined in (7)) and $\mathbf{n}(t_k)$, both with zero means. With the narrowband assumption for the stochastic signal model we assume that covariance matrix $\mathbf{P}$ of the random complex amplitudes do not vary with time and is given by

$$\mathbf{P} = E\left[\mathbf{s}(t_k)\mathbf{s}^H(t_k)\right] = E\left[\mathbf{a}(t)\mathbf{a}^H(t)\right] = E\left[\mathbf{a}\mathbf{a}^H\right].$$

Thus the array vector for the k^{th} snapshot has zero mean and covariance $\mathbf{D}(\boldsymbol{\theta})\mathbf{P}\mathbf{D}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}$ respectively. Furthermore since the signal amplitudes and noise samples are independent from snapshot to snapshot, the joint PDF of the array vectors $\mathbf{y}(t_k)$ for $k = 1, \ldots, N$, is given by

$$p(\mathbf{x}; \boldsymbol{\theta}, \mathbf{P}) = (\pi)^{-NP} \frac{1}{|\mathbf{D}(\boldsymbol{\theta})\mathbf{P}\mathbf{D}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I})|} \exp\left[-\sum_{k=1}^{N} \mathbf{y}^H(t_k)(\mathbf{D}(\boldsymbol{\theta})\mathbf{P}\mathbf{D}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I})^{-1}\mathbf{y}(t_k)\right] \tag{12}$$

where $\mathbf{x}$ is defined as

$$\mathbf{x} = \left[\mathbf{y}^T(t_1) \ \ldots \ \mathbf{y}^T(t_N)\right]^T$$

Thus the log likelihood function $L(\boldsymbol{\theta}, \mathbf{P})$, with unnecessary terms discarded is given, by

$$L(\boldsymbol{\theta}, \mathbf{P}) = -\ln\left|(\mathbf{D}(\boldsymbol{\theta})\mathbf{P}\mathbf{D}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I})\right| - \sum_{k=1}^{N} \mathbf{y}^H(t_k)(\mathbf{D}(\boldsymbol{\theta})\mathbf{P}\mathbf{D}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I})^{-1}\mathbf{y}(t_k). \tag{13}$$

Unlike, the deterministic signal model with known waveforms [20], here the likelihood function is nonlinear with respect to the DOAs as well as with respect to all the elements of the matrix $\mathbf{P}$. Thus, direct maximization of $L(\boldsymbol{\theta}, \mathbf{P})$ for an unknown covariance matrix requires a maximization over $(M^2 + M)$ real parameters. This is because there are M unknown θs and M^2 unknown parameters in the covariance matrix $\mathbf{P}$. Even though this does not belong to the class of partial linear least squares problem, where the parameter estimation is decoupled, it was shown in [6] that the estimation of the DOA vector $\boldsymbol{\theta}$ and the elements of the covariance matrix $\mathbf{P}$ is decoupled. The result of [6] is that the ML estimate of the DOA vector $\boldsymbol{\theta}$ is obtained by a maximization of a function over only the DOA angle θ_is. In particular it was shown that $\hat{\boldsymbol{\theta}}_{MLE}$ is obtained by maximizing the compressed likelihood function $L_c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, where $L_c(\boldsymbol{\theta})$ is given by (for proof, see [6]),

$$L_c(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\text{Trace}\left[\mathbf{G}(\boldsymbol{\theta})\mathbf{S}\right] - \ln\left|\mathbf{G}(\boldsymbol{\theta})\mathbf{S}\mathbf{G}(\boldsymbol{\theta}) + \sigma^2(\mathbf{I}_P - \mathbf{G}(\boldsymbol{\theta}))\right| \tag{14}$$

where

$$\mathbf{S} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{y}(t_k)\mathbf{y}^H(t_k)$$

and

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{D}(\boldsymbol{\theta})\left[\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})\right]^{-1}\mathbf{D}^H(\boldsymbol{\theta})$$

is the orthogonal projection matrix. The compressed likelihood function $L_c(\boldsymbol{\theta})$ can be written in terms of the N array snapshots $\mathbf{y}(t_k), k = 1, \ldots, N$ as

$$L_c(\boldsymbol{\theta}) = \frac{1}{N\sigma^2}\left[\sum_{k=1}^{N} \mathbf{y}^H(t_k)\mathbf{G}(\boldsymbol{\theta})\mathbf{y}(t_k)\right] - \ln\left|\mathbf{G}(\boldsymbol{\theta})\mathbf{S}\mathbf{G}(\boldsymbol{\theta}) + \sigma^2(\mathbf{I}_P - \mathbf{G}(\boldsymbol{\theta}))\right|. \tag{15}$$

The ML estimate of the covariance matrix is given by,

$$\hat{\mathbf{P}}_{\text{mle}} = \left[\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})\right]^{-1} \mathbf{D}^H(\boldsymbol{\theta}) \left[\mathbf{S} - \sigma^2 \mathbf{I}_P\right] \mathbf{D}(\boldsymbol{\theta}) \left[\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})\right]^{-1}. \tag{16}$$

for known $\boldsymbol{\theta}$ and σ^2. If σ^2 is not known it needs to be replaced by its estimate which is obtained as the $P-M$ smallest eigen values of the matrix $\hat{\mathbf{R}}$, [26] given by

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{y}(t_k)\mathbf{y}^H(t_k)$$

4 Global Maximization of the Likelihood Function

The ML estimate of the DOAs $\boldsymbol{\theta}$ is obtained by maximizing the compressed likelihood function $L_c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Its direct and most straightforward implementation requires a multidimensional grid search whose computational complexity increases exponentially with the number of sources. There is an enormous literature on solving multidimensional optimization problems of this kind which avoid a grid search. Most of the approaches are iterative in nature and require a good initial guess of the multidimensional parameter vector. The need for having a good initial guess, furthermore requires an application of a suboptimal but computationally modest algorithm to be applied on the problem before applying the ML technique. Simulated annealing [7], EM Algorithm [5], IQML algorithm [3], are some of the techniques which have been used to solve this type of problem. However, all these approaches are iterative and thus global convergence is not guaranteed. It has been demonstrated by means of simulations in [5] that the EM algorithm fails to converge to the true optimum if the initial guess has a significant deviation from the true values.

To overcome the drawback of these iterative techniques, we aim to use a technique which is not iterative and hence does not require an initial guess of the DOA estimates. Pincus [21] showed that for such problems, it is possible to obtain a closed form solution for the parameter $\mathbf{x}$, that yields the *global* optimum. Based on the theorem given by Pincus, the $\mathbf{x}$ that yields the *global* maximum of $L_c(\mathbf{x})$, is given by

$$\hat{x}_i = \lim_{\rho\to\infty} \frac{\int \dots \int x_i \exp(\rho L_c(\mathbf{x}))d\mathbf{x}}{\int \dots \int \exp(\rho L_c(\mathbf{y}))d\mathbf{y}} \qquad i = 1, 2, \dots, M. \tag{17}$$

If we let $L'(\mathbf{x}) = \exp(\rho L_c(\mathbf{x}))$, and the normalized version of $L^{'}(\mathbf{x})$ be $\bar{L}'(\mathbf{x}) = \frac{\exp(\rho L_c(\mathbf{x}))}{\int \dots \int \exp(\rho L_c(\mathbf{y}))d\mathbf{y}}$, then the function $\bar{L}^{'}(\mathbf{x})$ is non-negative and has all the properties of a PDF, although strictly speaking it is not a PDF since $\mathbf{x}$ is not random. We term $\bar{L}^{'}(\mathbf{x})$ a pseudo-PDF in $\mathbf{x}$. With this definition, we define the optimal x_i which maximizes $L_c(\mathbf{x})$ in (14) as

$$\hat{x}_i = \int \dots \int x_i \bar{L}^{'}(\mathbf{x})d\mathbf{x} \qquad i = 1, \dots, M. \tag{18}$$

for some large value of ρ. Intuitively, as $\rho \to \infty$, the function $\bar{L}^{'}(\mathbf{x})$ becomes a multidimensional Dirac delta function centered at the location of the maximum of $L(\mathbf{x})$. Thus, (18) yields the location of the maximum.

Now, the optimal $\hat{x}_i$ requires an evaluation of an M-dimensional integral, which is difficult to implement in practice. However, since $\bar{L}'(\mathbf{x})$ is a pseudo-PDF, we can interpret $\hat{x}_i$ as the expected value of x_i, where the expectation is calculated with respect to the pseudo-PDF $\bar{L}^{'}(\mathbf{x})$. It has been shown that for this type of problem, Monte Carlo approximation techniques can achieve good results without using direct integration [22]. A straightforward Monte Carlo integration approximation can be defined as

$$\hat{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k \tag{19}$$

where $\mathbf{x}_k$ is the k^{th} realization of the vector $\mathbf{x}$ distributed according to $\bar{L}^{'}(\mathbf{x})$. Computing $\hat{\mathbf{x}}$ by (19) requires generation of $\mathbf{x} \sim \bar{L}^{'}(\mathbf{x})$. For the problem of interest in this paper, generation of the vector $\mathbf{x} \sim \bar{L}^{'}(\mathbf{x})$ may not be easy, as $\bar{L}^{'}(\mathbf{x})$ is a highly nonlinear function of $\mathbf{x}$. So even though direct integration can be bypassed by using (18), generation of $\mathbf{x} \sim \bar{L}^{'}(\mathbf{x})$ may again demand integration. As a result, we do not use (18) to compute $\hat{\mathbf{x}}$. Rather we use importance sampling [22], as described in the next section.

5 Importance Sampling

To compute a multidimensional integral of the type given in (18), importance sampling has been shown to be a powerful tool. The approach is based on the observation that integrals of the type $\int h(\mathbf{x})\bar{L}'(\mathbf{x})\mathbf{dx}$ can be expressed as

$$\int h(\mathbf{x})\bar{L}^{'}(\mathbf{x})\mathbf{dx} = \int h(\mathbf{x})\frac{\bar{L}^{'}(\mathbf{x})}{\bar{g}(\mathbf{x})}\bar{g}(\mathbf{x})d\mathbf{x} \tag{20}$$

where $\bar{g}(\mathbf{x})$ is assumed to possess all the properties of a PDF. Then, the right-hand-side of (20) can be expressed as the expected value of $h(\mathbf{x})\frac{\bar{L}^{'}(\mathbf{x})}{\bar{g}(\mathbf{x})}$, with respect to the pseudo-PDF $\bar{g}(\mathbf{x})$. The function $\bar{g}(\mathbf{x})$ is called the normalized importance function. Unlike $\bar{L}^{'}(\mathbf{x})$, which in general is a nonlinear function of $\mathbf{x}$, $\bar{g}(\mathbf{x})$ can be chosen to be some simple function of $\mathbf{x}$, so that realizations of $\mathbf{x}$ can be easily generated. Then, the value of the integral in (20) can be found by the Monte Carlo approximation

$$\frac{1}{K} \sum_{k=1}^{K} h(\mathbf{x}_k)\frac{\bar{L}^{'}(\mathbf{x}_k)}{\bar{g}(\mathbf{x}_k)} \tag{21}$$

where $\mathbf{x}_k$ is the k^{th} realization of the vector $\mathbf{x}$ distributed according to the pseudo-PDF $\bar{g}(\mathbf{x})$. The value of K needed for a good approximation depends on the choice of g. Typically, $\bar{g}(\mathbf{x})$ should be chosen similar to $\bar{L}(\mathbf{x})$, as this reduces the variance of the estimate given by (21). However, another important point to keep in mind when choosing $\bar{g}(\mathbf{x})$ is that it should be simple enough so that $\mathbf{x} \sim \bar{g}(\mathbf{x})$ can be easily generated [17], [18] . We explain in the next section, how to choose $\bar{g}(\mathbf{x})$ for the deterministic signal and the stochastic signal case.

6 Choice of Importance Function

The compressed likelihood function from (14) is a combination of two terms, unlike the deterministic signal case [1]. Thus, it is not possible to make the function separable in the DOAs, as in the deterministic case. However, Jaffer [6] has shown that the two terms in the compressed likelihood function defined in (15) can be expressed in terms of the positive eigenvalues $\lambda_1, \ldots, \lambda_M$ of the matrix $\mathbf{G}(\boldsymbol{\theta})\mathbf{S}\mathbf{G}(\boldsymbol{\theta})$, (with $\mathbf{G}(\boldsymbol{\theta})$ and $\mathbf{S}$ defined earlier in Section 3) as

$$\frac{1}{N\sigma^2} \sum_{k=1}^{N} \mathbf{y}^H(t_k)\mathbf{D}(\boldsymbol{\theta})(\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta}))^{-1}\mathbf{D}^H(\boldsymbol{\theta})\mathbf{y}(t_k) = \sum_{i=1}^{M} \frac{\lambda_i}{\sigma^2}. \tag{22}$$

and

$$\ln\left|\mathbf{G}(\boldsymbol{\theta})\mathbf{S}\mathbf{G}(\boldsymbol{\theta}) + \sigma^2(\mathbf{I}_P - \mathbf{G}(\boldsymbol{\theta}))\right| = \ln\left[(\sigma^2)^{P-M} \prod_{i=1}^{M} \lambda_i\right] = \ln\left[(\sigma^2)^{P} \prod_{i=1}^{M} \frac{\lambda_i}{\sigma^2}\right]. \tag{23}$$

respectively. (23) can be written as,

$$\ln\left|\mathbf{G}(\boldsymbol{\theta})\mathbf{S}\mathbf{G}(\boldsymbol{\theta}) + \sigma^2(\mathbf{I}_P - \mathbf{G}(\boldsymbol{\theta}))\right| = P\ln(\sigma^2) + \sum_{i=1}^{M} \ln(\frac{\lambda_i}{\sigma^2}). \tag{24}$$

From (22) and(24), whose left hand sides are the first and second term respectively of the compressed likelihood function defined in (14), $L_c(\boldsymbol{\theta})$ can be written in terms of the eigenvalues λ_is as

$$L_c(\boldsymbol{\theta}) = \left[\sum_{i=1}^{M} \frac{\lambda_i}{\sigma^2}\right] - \left[P\ln(\sigma^2) + \sum_{i=1}^{M} \ln(\frac{\lambda_i}{\sigma^2})\right]. \tag{25}$$

Since our aim is to choose an importance function $g(\boldsymbol{\theta})$ from which $\boldsymbol{\theta}$ can be generated easily, we choose $g(\boldsymbol{\theta})$ by considering only the first term of the compressed likelihood function and further approximating it by forcing $\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta}) = P\mathbf{I}$. Note that for large SNRs, and well separated θs, this is a close approximation to the actual compressed likelihood function. This can be observed by noting the first and second terms within brackets in the right hand side of (25). Since $\lambda_i >> \sigma^2$ and $[P\ln(\sigma^2)]$ is small for high SNRs, the contribution due to $\sum_{i=1}^{M} \frac{\lambda_i}{\sigma^2}$ is expected to be higher compared to $\left[P\ln(\sigma^2) + \sum_{i=1}^{M} \ln(\frac{\lambda_i}{\sigma^2})\right]$. Furthermore for well separated θs, the first term of the compressed likelihood function in (14), is Trace $[\mathbf{G}(\boldsymbol{\theta})\mathbf{S}] = \frac{1}{\sigma^2}\sum_{k=1}^{N} \mathbf{y}^H(t_k)\mathbf{D}(\boldsymbol{\theta})\left[\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})\right]^{-1}\mathbf{D}^H(\boldsymbol{\theta})\mathbf{y}(t_k) \approx$

$\frac{1}{\sigma^2}\sum_{k=1}^{N}\mathbf{y}^H(t_k)\mathbf{D}(\boldsymbol{\theta})(\frac{\mathbf{I}}{P})\mathbf{D}^H(\boldsymbol{\theta})\mathbf{y}(t_k)$, because $\left[\mathbf{D}^H(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})\right]^{-1} \approx \frac{\mathbf{I}}{\mathbf{P}}$. We choose the importance function for the stochastic signal case from (22), as

$$g(\boldsymbol{\theta}) = \exp\left(\rho_1 \frac{1}{N\sigma^2}\sum_{k=1}^{N}\mathbf{y}^H(t_k)\mathbf{D}(\boldsymbol{\theta})\frac{\mathbf{I}}{P}\mathbf{D}^H(\boldsymbol{\theta})\mathbf{y}(t_k)\right). \tag{26}$$

Rewriting $g(\boldsymbol{\theta})$ in the same way as for the deterministic signal case,

$$g(\boldsymbol{\theta}) = \prod_{i=1}^{M}\exp(\rho_1 I_y(\theta_i)) \tag{27}$$

where

$$I_y(\theta_i) = \frac{1}{\sigma^2 N}\sum_{k=1}^{N}\left[\frac{1}{P}\left|\sum_{l=0}^{P-1} y_l(t_k)\exp(-j\pi l\cos(\theta_i))\right|^2\right]. \tag{28}$$

Note that $I_y(\theta)$ is a time averaged periodogram of the individual spatial array snapshots The choice of the number ρ_1 in (26) is highly problem specific. We discuss that in Section 7. We emphasize that *no alterations* are made to the *actual modified likelihood function* $L_c(\boldsymbol{\theta})$. We retain both the terms in it when global optimization theorem is being used with importance sampling to compute the DOA estimates.

6.1 Estimation of DOAs

The global optimization theorem demands evaluation of the means of $\theta_1, \ldots, \theta_M$, distributed according to the normalized modified compressed likelihood function. Since the DOAs $\theta_1, \ldots, \theta_M$ are bounded from below and above, they possess the properties of a circular variable rather than a linear random variable. If the mean were to be evaluated directly as

$$\hat{\theta}_{i,\text{mle}} = \int \ldots \int \theta_i \bar{L}(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{29}$$

then the estimates obtained would be biased [12], especially at low SNRs and/or for short data records. Hence, we use the circular mean, instead of linear mean (29), to estimate the DOA. The circular mean is defined as

$$\hat{\theta}_{i,\text{mle}} = \frac{1}{2\pi}\angle \int \ldots \int \exp(j2\pi\theta_i)\bar{L}(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{30}$$

This amounts to computing the angle of the mean of $\exp(j2\pi\theta_i)$ with respect to the modified likelihood function. The key idea in defining a circular mean is to average position vectors. Hence, if $\alpha_1, \alpha_2, \ldots, \alpha_R$ are R realizations of a random point $\exp(j\alpha)$ on the circumference of a circle of unit radius, then the circular sample mean [19] of the data is defined as

$$\bar{\alpha} = \angle \frac{1}{R}\sum_{k=1}^{R}\exp(j\alpha_k). \tag{31}$$

The use of (31) alleviates the estimator bias. The difficulty of using the linear mean had also been remarked upon by Lovell [14]. In this problem of DOA estimation of the angles $\theta_i, i = 1, \ldots, M$, we use the circular mean concept to obtain the MLE of the function of the angle ϕ_i, given by $\phi_i = \frac{1}{2}\cos(\theta_i)$ using (21) and (31) and then use the inverse transformation $\theta_i = \cos^{-1}(2\phi_i)$ to obtain the MLE of the angle $\theta_i, i = 1, \ldots, M$. This is valid as a consequence of the invariance properties of maximum likelihood estimates. Furthermore it should be noted that $-0.5 \leq \phi_i \leq 0.5$, as the DOA angles have the range $0 \leq \theta_i \leq \pi, i = 1, \ldots, M$. Thus, $\hat{\phi}_{i,\text{mle}}$ is defined from (21) and (31) as

$$\hat{\phi}_{i,\text{mle}} = \frac{1}{2\pi} \angle \frac{1}{R} \sum_{k=1}^{R} \frac{\bar{L}(\boldsymbol{\phi}_k)}{\bar{g}(\boldsymbol{\phi}_k)} \exp(j2\pi[\boldsymbol{\phi}_k]_i) \tag{32}$$

for $i = 1, \ldots, M$, where $\boldsymbol{\phi}_k$ is the k^{th} realization of the transformed DOA angle vector, and $[\boldsymbol{\phi}_k]_i$ refers to the i^{th} component. Note that since we need only find the angle of the complex quantity in (32), an equivalent estimator is

$$\hat{\phi}_{i,\text{mle}} = \frac{1}{2\pi} \angle \frac{1}{R} \sum_{k=1}^{R} \frac{L'(\boldsymbol{\phi}_k)}{g(\boldsymbol{\phi}_k)} \exp(j2\pi[\boldsymbol{\phi}_k]_i) \tag{33}$$

or finally

$$\hat{\phi}_{i,\text{mle}} = \frac{1}{2\pi} \angle \frac{1}{R} \sum_{k=1}^{R} w(\boldsymbol{\phi}_k) \exp(j2\pi[\boldsymbol{\phi}_k]_i) \tag{34}$$

where

$$w(\boldsymbol{\phi}) = \frac{L'(\boldsymbol{\phi})}{g(\boldsymbol{\phi})}. \tag{35}$$

This observation is quite important in that it simplifies the computation greatly. We no longer need to find the normalization constants $\int L'(\boldsymbol{\phi})d\boldsymbol{\phi}$ and $\int g(\boldsymbol{\phi})d\boldsymbol{\phi}$ in computing $\bar{L}(\boldsymbol{\phi})$ and $\bar{g}(\boldsymbol{\phi})$. Equivalence of (32) and (34) has been possible because ϕ_i is bounded from above and below for all i. This equivalence has further advantages in carrying out computations, which we now discuss. Even though (34) and (35) are the expressions for computing $\hat{\phi}_{i,\text{mle}}$ and $w(\boldsymbol{\phi})$ respectively, it is important to note that this may result in computation difficulties because both the numerator and denominator of right hand side of (35) are exponentials. To overcome this difficulty, $w(\boldsymbol{\theta})$ is computed as

$$w(\boldsymbol{\phi}) = \frac{L'(\boldsymbol{\phi})}{g(\boldsymbol{\phi})} = \exp(\rho L_c(\boldsymbol{\phi}) - \rho_1 \sum_{i=1}^{M} I(\phi_i)) \tag{36}$$

since

$$L'(\boldsymbol{\phi}) = \exp\left(\rho L_c(\boldsymbol{\phi})\right)$$

and

$$g(\boldsymbol{\phi}) = \exp\left(\sum_{i=1}^{M} \rho_1 I(\phi_i)\right)$$

where $I(\phi_i) = I_y(\theta_i)$ defined in (28). Furthermore, if we define $w'(\boldsymbol{\phi}_k)$ as

$$w'(\boldsymbol{\phi}_k) = \exp\left[\rho L_c(\boldsymbol{\phi}_k) - \sum_{i=1}^{M} \rho_1 I([\boldsymbol{\phi}_k]_i) - \max_{1\le k\le R}\left(\exp(\rho L_c(\boldsymbol{\phi}_k) - \sum_{i=1}^{M} \rho_1 I([\boldsymbol{\phi}_k]_i)\right)\right] \tag{37}$$

then $w'(\boldsymbol{\phi}_k)$ is a scalar multiple of $w(\boldsymbol{\phi}_k)$, and both are real quantities. The subscript k in (37) refers to the k^{th} realization of the generated vector $\boldsymbol{\phi}$. Note that $w'(\boldsymbol{\phi}_k)$ in (37) can be obtained only after all the R realizations of $w(\boldsymbol{\phi}_k)$ have been obtained. Thus we can replace $w(\boldsymbol{\phi})$ in (34) by $w'(\boldsymbol{\phi})$ defined in (37) so that

$$\hat{\phi}_{i,\mathrm{mle}} = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R} w'(\boldsymbol{\phi}_k)\exp(j2\pi[\boldsymbol{\phi}_k]_i). \tag{38}$$

(38) is equivalent to (34) , because the $\angle$ operator is invariant to multiplication of an argument by a positive constant. It is (38) which removes the difficulty of handling extremely large numbers which would have been involved if $\hat{\phi}_{i,\mathrm{mle}}$ were computed using (34).

7 Summary of Steps of Implementation

1. From the array snapshots, $\mathbf{y}(t_k)$ for $k = 1, \dots, N$, evaluate the time averaged periodogram of the individual array snapshots $I_y(\boldsymbol{\theta})$ using (28) to obtain the normalized function

$$g(\phi_i) \approx \frac{\exp(\rho_1 I(\phi_i))}{\sum_{i=1}^{L}\exp(\rho_1 I(\phi_i))}$$

   where L is the total number of points at which the spatial periodogram is evaluated. For the stochastic signal case obtain $g(\phi_i)$ in the same way except that $I(\phi_i)$ is given by (28).

2. Generate the vector $\boldsymbol{\phi} \sim \bar{g}(\boldsymbol{\phi})$, where $\phi_i = \frac{1}{2}\cos(\theta_i), i = 1, \dots, M$ and $g(\boldsymbol{\theta})$ has been defined in (26). To do so, generate $\phi_1 \sim g(\phi)$ obtained in the previous step. Once ϕ_1 is obtained, generate $\phi_2 \sim g(\phi)$, with the condition that it is not the same as ϕ_1. Continue generating $\phi_3, \dots, \phi_M$ in a similar way with the condition that they are all distinct. For a discussion on the implementation details of this step, the reader is referred to [15].

3. Repeat step 2 R times to obtain R realizations of the vector $\boldsymbol{\phi}$.

4. Evaluate the DOA estimate from the estimate of $\hat{\phi}_{i,\mathrm{mle}}$ as $\hat{\theta}_{i,\mathrm{mle}} = \cos^{-1}(2\hat{\phi}_{i,\mathrm{mle}})$ for $i = 1, \dots, M$ using

$$\hat{\phi}_{i,\mathrm{mle}} = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R} w'(\boldsymbol{\phi}_k)\exp\left(j2\pi[\boldsymbol{\phi}_k]_i\right) \tag{39}$$

   where $w'(\boldsymbol{\phi}_k)$ has been defined in (37).

In (37), ρ should be chosen very high. ρ_1 should be chosen judiciously in order to make R small in (39). Recall that $g(\boldsymbol{\phi})$ should be chosen as a close approximation to $L_c(\boldsymbol{\phi})$. Since $I(\phi)$ is expected to produce biased estimates of the DOAs, which need to be alleviated by the weighting $w(\boldsymbol{\phi})$, it makes sense not to increase the bias by choosing ρ_1 too high. On the other hand choosing ρ_1 too low may result in spurious peaks, as the peaks due to noise may also result in a contribution. Although the final estimates are only slightly affected by the choice of ρ_1, in order to make R as low as possible and hence minimize the computational burden, an appropriate choice of ρ_1 is important.

8 Simulation Results

In order to demonstrate the performance of the method we now describe the simulation results. Since the major interest in ML methods is because of the superior performance under threshold conditions where many of the suboptimal techniques fail, we carry out simulations under threshold conditions. Thus we choose a low SNR, short data records (small number of snapshots), and closely spaced sources in our simulations and show its superiority compared to MUSIC. Since another implementation of the ML technique (using the EM algorithm) was given in [5], we compare our results with those of [5]. MUSIC is also applied in order to demonstrate its failure under threshold conditions, where the ML technique appears to be the only resort.

Our simulation example is the same as that in [5]. We chose the same simulation example as [5] as the conditions under which the simulations in [5] were performed were threshold. By threshold conditions we mean that the SNR and the number of snapshots are low, under which the suboptimal techniques like MUSIC fail. There are 9 sensors, arranged in a uniform linear array with $\frac{\lambda}{2}$ spacings, and there are $M = 2$ directional narrowband sources. For the first experiment, the sources are located at $\theta_1 = 45^o$ and $\theta_2 = 60^o$. To make a one to one comparison of quantitative results with that using the EM algorithm reported in [5], we perform 100 Monte Carlo estimations of the DOAs. The scatter plots for the angle estimates are shown in Figure 2, for 100 realizations. It can be observed that most of the estimates are centered around the true DOAs. In Figure 3, the scatter plots for this example using MUSIC is shown. It is obvious from Figure 3 that MUSIC fails to estimate 1 source on about 11 occasions. This is because of resolution failure in the MUSIC spectrum in those realizations. The second angle estimate is generated from some smaller peak in the MUSIC spectrum well separated from the true source location. We also describe the performance quantitatively by evaluating the bias and root mean square error of the estimates. The root mean square error and bias obtained using the proposed technique are shown in Table 1. Table 2 shows the bias and root mean square error using the EM algorithm, for the same example with three initial guesses. It can be observed from Table 1 and 2 that both EM algorithm [5] and the proposed method have similar performance for this example.

In Figures 4 and 5, and Tables 3 and 4, we present performance results for the case as in the example just considered except the fact that the sources are now more closely spaced. For this example, we have the same SNR, and same number of snapshots , but the sources are located at $\theta_1 = 48^o$ and $\theta_2 = 55^o$. From the scatter plots shown in Figure 5, it is obvious that MUSIC fails to estimate the DOAs in most of the Monte Carlo runs for this example. In Tables 3 and 4, we give the quantitative performance for the proposed method and the EM algorithm respectively. It can be observed that the root mean square error obtained using our technique is the nearly the same as that obtained using the EM algorithm if the initial guess of the angle estimates have a 10^o deviation from their respective true values. Our mean square error values are slightly higher than that using EM algorithm if the initial guess using the EM algorithm has only 1^o deviation from the respective true values. However, it is more justifiable to compare the results obtained using the proposed technique with the EM algorithm for the cases when initial guess for the angles have a larger deviation from the true values, as under these threshold conditions, it is difficult to get a good initial guess by use of some other suboptimal algorithm. It can be noted that the root mean square errors for initial guesses having deviations 5^o and 10^o are almost the same as that of the proposed technique. This is because both are implementations of the ML technique. Also it can be observed from Table 4, that there is a bias in the estimates using the EM algorithm whereas the bias using the proposed method is negligible.

In all the simulations we choose $\rho_1 = 0.02$ and observe that the estimates do not change with an increase in ρ for $\rho = 800$ and larger. Although we obtain identical performance by changing ρ_1 and ρ, the number of importance sampling realizations R was found to be the least for choice of ρ_1 in the range $0.01 \leq \rho_1 \leq 0.06$. For $0.06 < \rho_1 \leq 0.2$, R had to be increased to 1200.

Next we consider a simulation example similar to that in [1]. The purpose of doing this is to compare the proposed technique to that in [1]. Although MODE/WSF [1] is not an exact ML implementation, but rather a large sample equivalent of the ML technique, we compare the performance as MODE exhibits results similar to ML method for uncorrelated sources. Table 5, shows the results for the proposed method, MODE and MUSIC for the simulation example in [1]. In this simulation there are five sensors placed over an uniformly spaced array with sensors at half wavelength spacing from each other. The SNR is 0 dB and the number of snapshots is 256. The two sources have bearings of 10^o and 25^o relative to the broadside. The root mean square and the bias have been calculated using 32 runs. Table 6 shows the results for the same example except that the number of snapshots is 64. From table 5 and 6, for the considered simulation example, it can be observed that MODE and the proposed method perform equally well. However MODE is computationally even less burdensome. This is discussed in the next section.

9 Conclusions and Future Directions

We have developed a computationally modest technique to implement a ML technique to estimate DOAs when the sources are stochastic. It is important to note that our technique does not require any initial guess of the angle estimates but still performs as well as the EM implementation of the ML method. Furthermore, convergence is not an issue at all using the proposed method, as it is not iterative. The number of importance sampling realizations R required in (39) for satisfactory performance is as low as 500, which means it requires only 500 evaluations of the compressed likelihood functions. This is far less compared to a grid search which even in the absence of noise requires at least 10000 evaluations of the likelihood function when there are 2 sources. The number of evaluations using a grid search also grows exponentially with an increase in the number of the sources, whereas this is not the case using the proposed technique. For a more detailed discussion on this issue the reader is referred to [15], which addresses the frequency estimation problem with unknown deterministic amplitudes.

Although the method performs well when sources are closely spaced in low SNR environment where suboptimal methods like MUSIC fail, MODE/WSF [1], [25], [24] is the most promising method reported so far in terms of performance and computational complexity. This is because MODE has been reported to be computationally comparable to that of MUSIC. Even though the MODE is in theory, a large sample approximation of the ML method, it has been demonstrated by means of simulations [1], to perform as well as the ML method when sources are uncorrelated and better than the ML method when the sources are correlated. Thus MODE is better than the proposed method with respect to computational complexity.

The performance of the method when the sources become correlated needs to be determined. This is currently under investigation. The global optimization of the MODE cost function using the proposed method may also be pursued as MODE has been proved to be statistically more efficient than the ML method in large samples [1].

References

[1] Petre Stoica and K. Sharman, "Maximum likelihood methods for direction of arrival estimation," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 38, pp. 1132-1143, July 1990.

[2] Petre Stoica and A. Nehorai, " Performance study of conditional and unconditional direction of arrival estimation, for direction of arrival estimation," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 38, pp. 1783-1795, October 1990.

[3] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise", *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 59, no. 12, October 1986, 1081-1089.

[4] Jian Li, R.T. Compton, "Maximum likelihood angle estimation for signals with known waveforms *IEEE Trans. Signal Processing*, vol.41, no.9, 1993, pp. 2850-2861.

[5] M.I. Miller, D.R. Fuhrmann, "Maximum likelihood narrow-band direction finding and the EM Algorithm, *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 38, no. 9, pp. 1560-1577, 1990.

[6] A.G. Jaffer, "Maximum likelihood direction finding for stochastic sources : A separable solution *IEEE , International Conference on Acoust., Speech, and Signal Processing*, pp. 2893-2897, 1988.

[7] K. Sharman, "Maximum likelihood parameter estimation by simulated annealing *IEEE , International Conference on Acoust., Speech, and Signal Processing*, pp. 2741-2744, 1988.

[8] I. Ziskind, M.Wax, Maximum likelihood localization of multiple sources by alternating projection, *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553-1560, 1988.

[9] S.S. Reddi, Multiple source location - A digital approach, " *IEEE Trans. Aerosp. Electron. Systems*, pp. 95-105, 1979.

[10] J. Capon, " High resolution frequency-wavenumber spectrum analysis, *Proc. IEEE*, vol. 57, pp. 1408-1418, 1969.

[11] S. Kay, "Fundamentals of Statistical Signal Processing," , Estimation Theory, Prentice Hall, Englewood Cliffs, NJ, 1993.

[12] S.M. Kay, "Comments on Frequency Estimation by Linear Prediction", *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp.198-199, April 1979.

[13] D.H. Johnson, "The application of spectral estimation methods to bearing estimation problems", *Proc. IEEE*, vol. 70, pp. 1018-1028, 1982.

[14] B.C. Lovell, P.J. Kootsookos, R.C. Williamson, "The Circular Nature of Discrete-Time Frequency Estimates", *Pro. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May, 1991, pp. 3369–3372.

[15] S. Kay, S. Saha, "Mean Likelihood Frequency Estimation", *IEEE Trans. Signal Processing*, July 2000.

[16] S. Kay, "Fundamentals of Statistical Signal Processing," , Detection Theory, Prentice Hall, Englewood Cliffs, NJ, 1998.

[17] Rubinstein, R.Y.,*Monte Carlo optimization, simulation, and sensitivity of queuing networks*, Wiley series in probability and mathematical statistics, 1986.

[18] M.H. Kalos, P.A. Whitlock, , *Monte Carlo Methods*, J. Wiley and Sons, 1986.

[19] K.V. Mardia, *Statistics of Directional Data*, Academic Press, London and New York, 1972.

[20] Supratim Saha, *Optimal Estimation of Linear and Nonlinear Parameters in General Linear Models by Importance Sampling*, Doctoral Dissertation, University of Rhode Island, April 2001.

[21] M. Pincus, "A Closed Form Solution for Certain Programming Problems", *Oper. Research*, pp. 690-694, 1962.

[22] L. Stewart, "Bayesian Analysis using Monte Carlo Integration: a Powerful Methodology for Handling Some Difficult Problems", *American Statistician*, pp. 195–200, 1983.

[23] A.F.M. Smith, A. Gelfand, "Bayesian Statistics without Tears: A Sampling-Resampling Framework", *American Statistical Association*, 1992

[24] Petre Stoica, K.C. Sharman, " Novel eigenanalysis method for direction estimation", *IEE Proceedings, pt. F*, vol. 137, pp. 19-26, february 1990.

[25] M. Viberg and B. Otterson, " Sensor array processing based on subspace fitting, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 39, pp.1110-1121, May 1991.

[26] Joakim Sorelius, "Subspace-Based parameter estimation problems in signal processing ", Thesis, Uppsala University, Systems and Control, department of information technology.

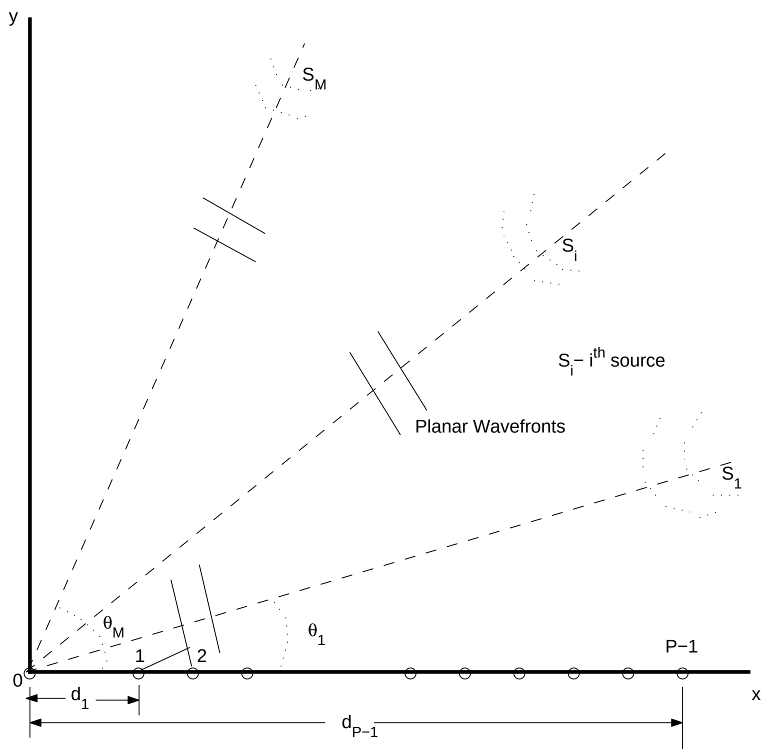

Figure 1: Geometry of the array for DOA estimation

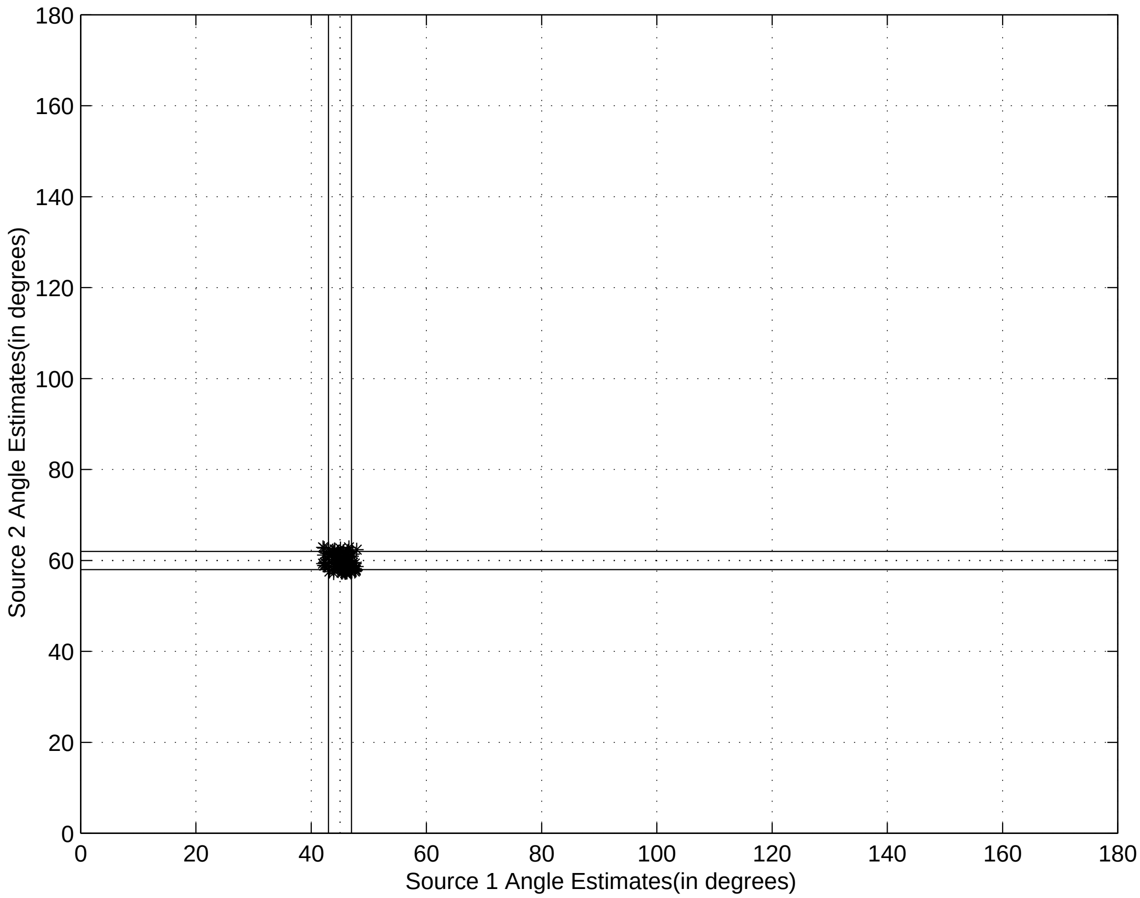

Figure 2: Scatter plots of angle estimates, stochastic signal model, using the proposed method $\theta_1 = 45^o, \theta_2 = 60^o$

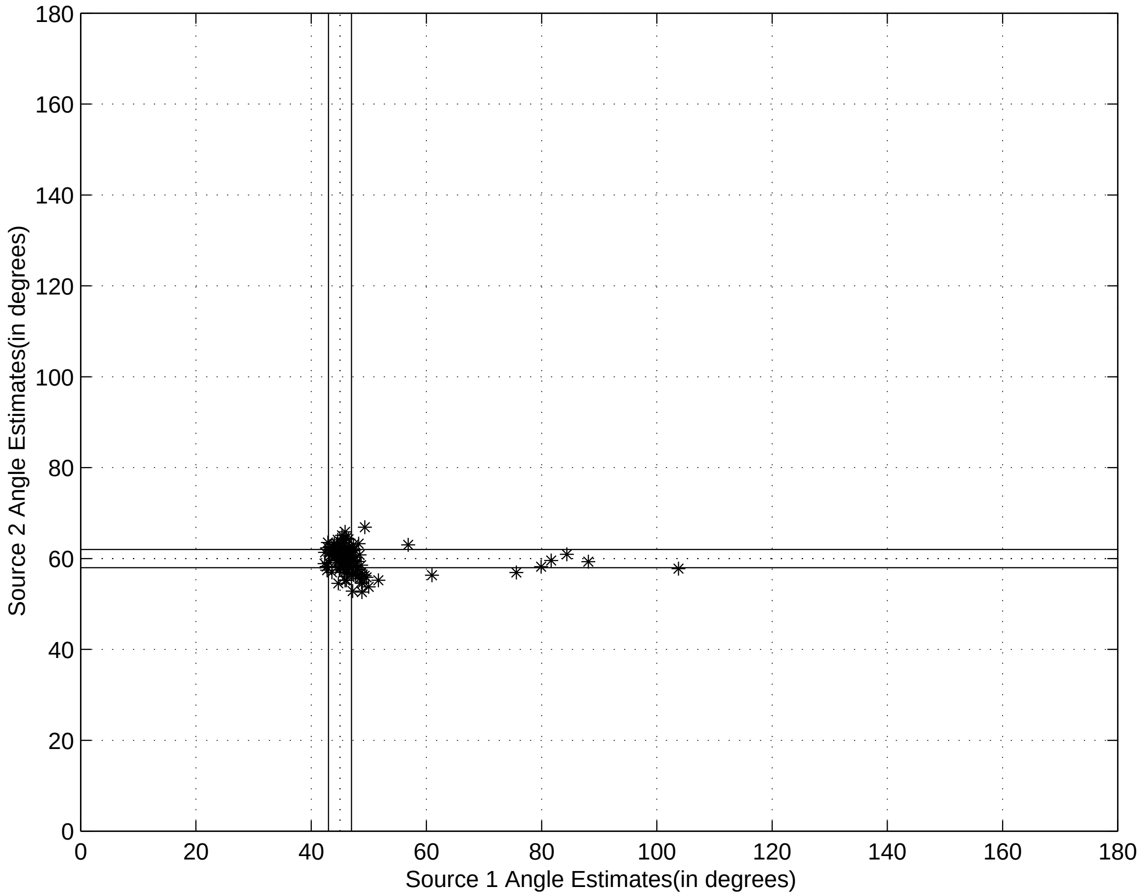

Figure 3: Scatter plots of angle estimates, stochastic signal model, using MUSIC $\theta_1 = 45^o, \theta_2 = 60^o$

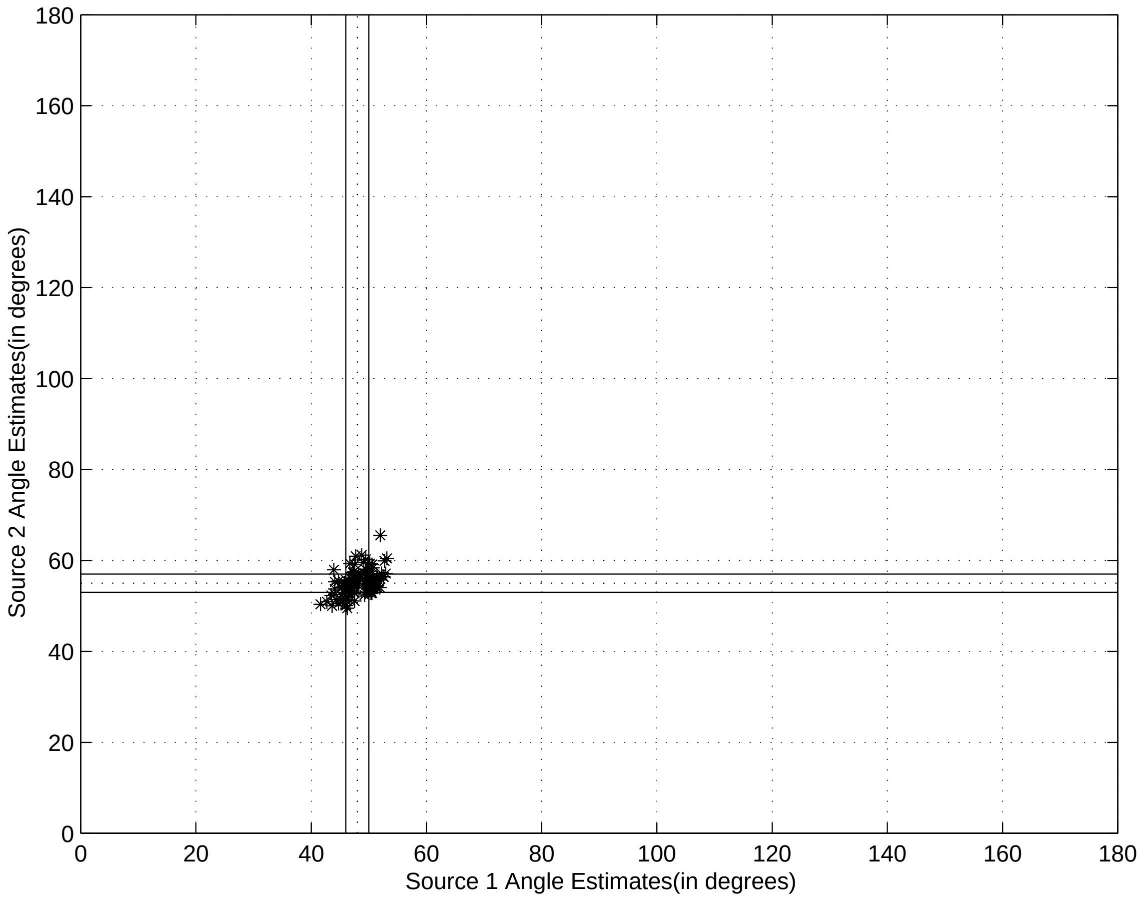

Figure 4: Scatter plots of angle estimates, stochastic signal model, using the proposed method $\theta_1 = 48^o, \theta_2 = 55^o$

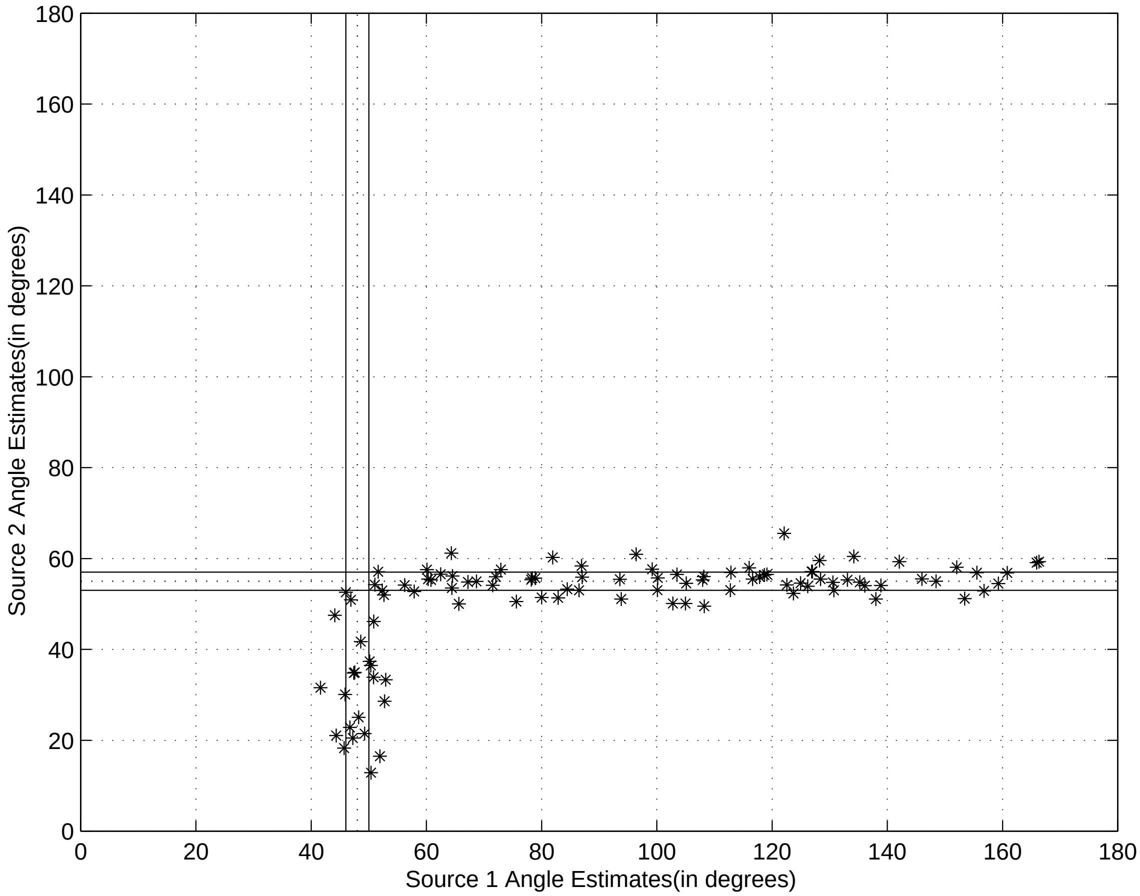

Figure 5: Scatter plots of angle estimates, stochastic signal model, using MUSIC $\theta_1 = 48^o, \theta_2 = 55^o$

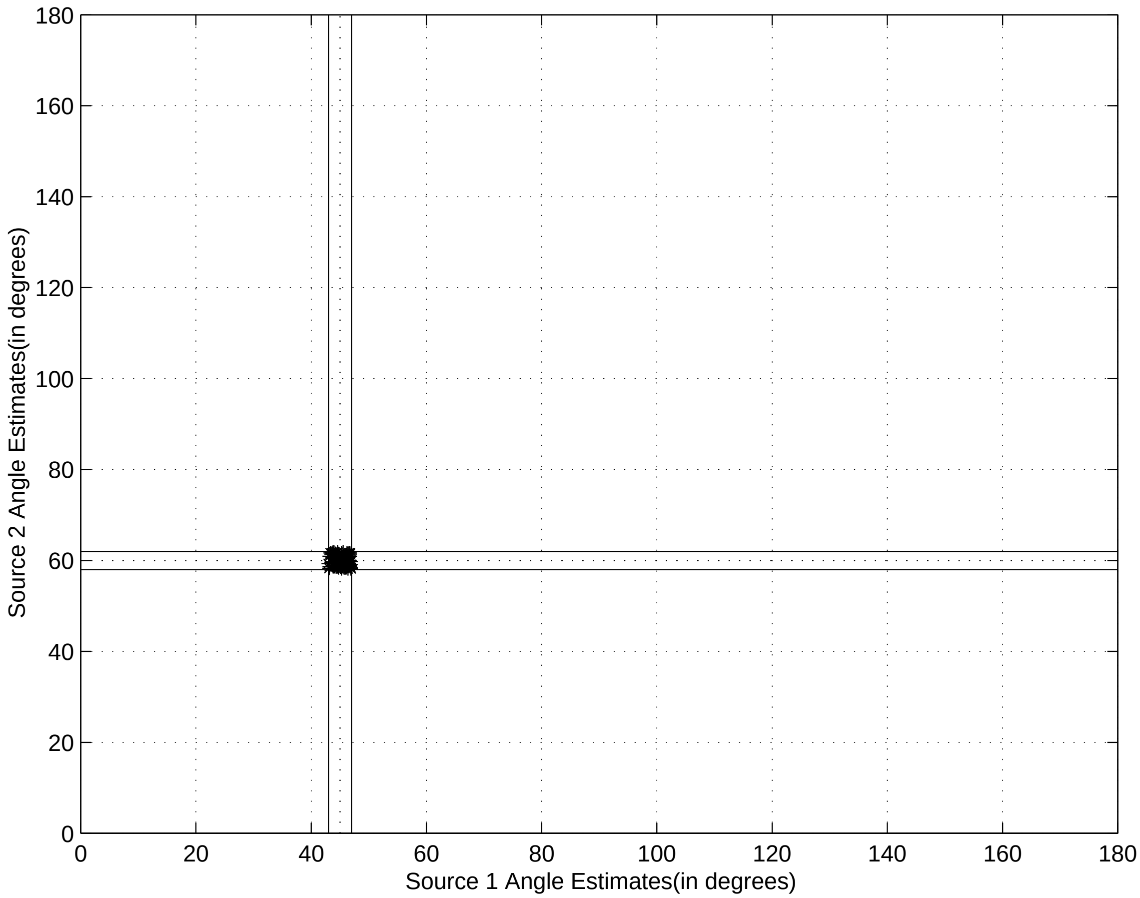

Figure 6: Scatter plots of angle estimates, deterministic signal model, using the proposed method $\theta_1 = 45^o, \theta_2 = 60^o$

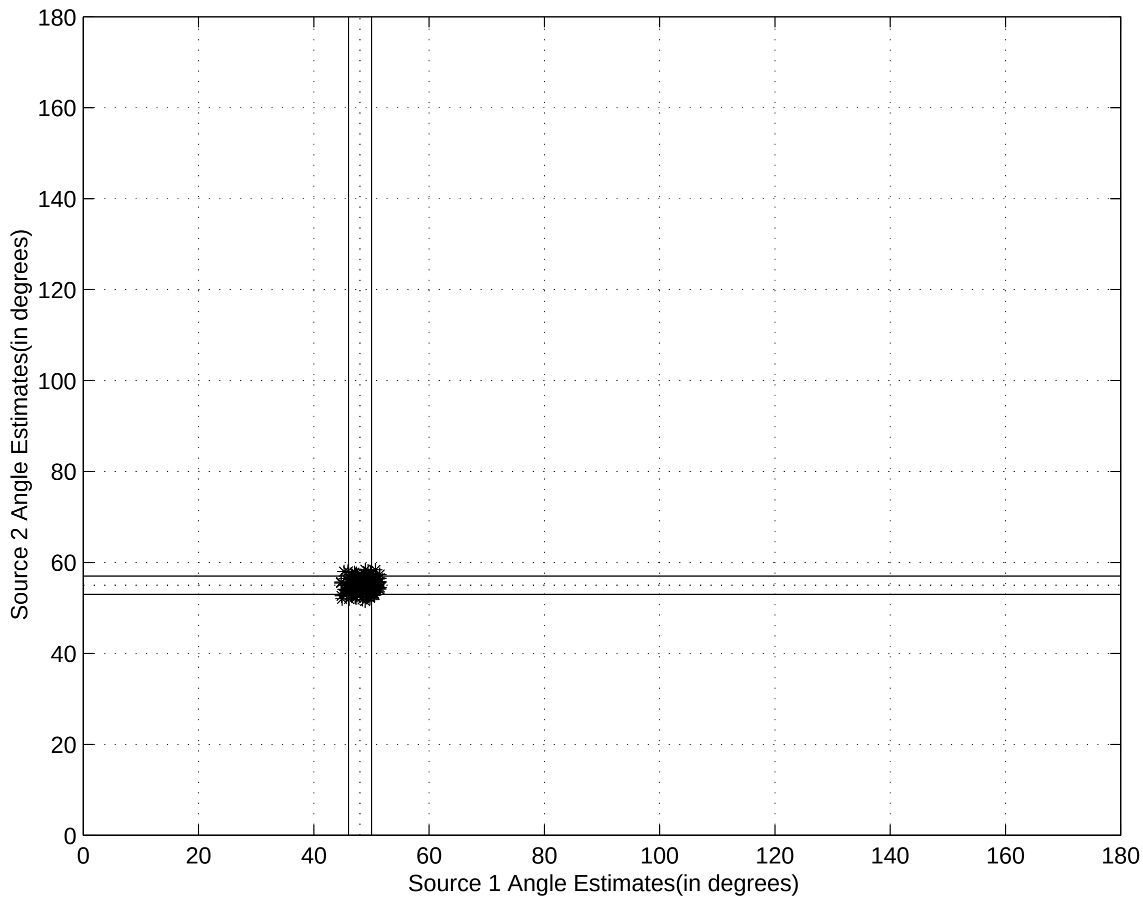

Figure 7: Scatter plots of angle estimates, deterministic signal model, using the proposed method $\theta_1 = 48^o, \theta_2 = 55^o$

Bias(θ_1)	Bias(θ_2)	RMS error (θ_1)	RMS error (θ_2)
0.23^o	-0.18^o	1.32^o	1.256^o

Table 1: Bias and root mean square errors using the proposed technique; $\theta_1 = 45^o, \theta_2 = 60^o$

Initial Guess ($\delta\theta$)	Bias(θ_1)	Bias(θ_2)	RMS error (θ_1)	RMS error (θ_2)
1^o	$-.6^o$	$+0.6$	1.22^o	1.22^o
5^o	$-.7^o$	$+0.7$	1.3^o	1.36^o
10^o	$-.7^o$	$+0.7$	1.3^o	1.3^o

Table 2: Bias and root mean square errors using EM Algorithm; $\theta_1 = 45^o, \theta_2 = 60^o$

Bias(θ_1)	Bias(θ_2)	RMS error (θ_1)	RMS error (θ_2)
0.36^o	-0.33^o	2.6^o	2.7^o

Table 3: Bias and root mean square errors using the proposed technique; $\theta_1 = 48^o, \theta_2 = 55^o$

Initial Guess ($\delta\theta$)	Bias(θ_1)	Bias(θ_2)	RMS error (θ_1)	RMS error (θ_2)
1^o	-1.2^o	$+0.7^o$	2.6^o	2.1^o
5^o	-1.1^o	$+0.9^o$	2.5^o	2.4^o
10^o	0.1^o	$+1.4^o$	2.6^o	2.8^o

Table 4: Bias and root mean square errors using EM Algorithm; $\theta_1 = 48^o, \theta_2 = 55^o$

Method	RMSE	Bias
Proposed	.471	0.016
MODE	0.472	0.015
MUSIC	0.477	0.014

Table 5: Bias and root mean square errors using the proposed method, MODE and MUSIC, with 5 sensors, source 1 at 10^0 and source 2 at 25^0, Number of snapshots =256.

Method	RMSE	Bias
Proposed	.911	0.01
MODE	0.908	0.011
MUSIC	0.933	0.010

Table 6: Bias and root mean square errors using the proposed method, MODE and MUSIC, with 5 sensors, source 1 at 10^0 and source 2 at 25^0, Number of snapshots =32.