

A Maximum Likelihood Angle-Doppler Estimator using Importance Sampling

Huigang Wang, *Member, IEEE*

Steven Kay, Fellow, IEEE,

voice: 401-874-5804

fax: 401-782-6422

Abstract

A new joint angle-Doppler maximum likelihood estimator based on importance sampling (IS) is proposed. The IS method allows one to compute the maximum likelihood estimator in a computationally efficient manner. It is based upon generating random variates using an importance function which approximates the compressed likelihood function. The performance is very close to the Cramer-Rao Lower Bound (CRLB). The choice of the algorithm parameters, which will affect the estimation performance, is also addressed in this paper. With a reasonable parameter choice, even the angles/Dopplers for closely spaced sources can be accurately estimated, whereas conventional subspace methods fail. Comparison with some suboptimal methods demonstrates that the IS method produces better performance at low SNR and/or small snapshots.

I. INTRODUCTION

In many signal processing problems such as are encountered in radar and sonar, it is desired to estimate parameters of multiple moving targets from the observed data. The direction of arrival (DOA) and the Doppler of the targets are two important parameters used to identify the target. This issue has been studied extensively in recent years [1], [2]. Some techniques such as high-resolution DOA estimation and frequency estimation, can efficiently estimate DOA and Doppler individually, but there have been few results addressing the joint estimation problem. As a result, a maximum likelihood angle-frequency estimator that can be efficiently implemented is desired, and is the subject of this paper.

Huigang Wang (wanghg74@nwpu.edu.cn) is with the College of Marine, Northwestern Polytechnical University, Xian, China, 710072

Steven Kay (kay@ele.uri.edu) is with Dept. of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881

It is well known that the maximum likelihood estimator (MLE) is an asymptotically efficient estimator [11]. In practice, its performance even for short data records nearly always outperforms other estimators. However, for the problem at hand the analytical expression for the likelihood function is a complicated nonlinear function of the parameters of interest and thus, it is computationally prohibitive to find the global maximum. Some suboptimal methods or simplified methods have been proposed to solve this problem in two different ways. The first approach is to decouple DOA and Doppler by replacing the 2-D optimization problem with two sequential 1-D optimization problems [3]. In this way, one set of parameters is fixed and the solution for the other set is found using a 1-D optimization algorithm. Then, the solution for the other parameter set is solved in the same way. Repeating this process results in an iterative approach, which, unfortunately cannot be shown to converge and even when convergence is obtained, is often only to a local maximum. The performance of these type of estimators will only attain the 2-D CRLB when the number of array elements and/or number of snapshots is large. As examples, in [6] a simple and efficient tree-structured frequency-space-frequency(FSF) MUSIC/ESPRIT algorithm, where two 1-D frequency-MUSIC/ESPRIT and a 1-D space-MUSIC/ESPRIT algorithm, accompanied by a temporal or spatial filtering process, approximates the 2-D nonlinear optimization required to determine the MLE. The FSF-MUSIC/ESPRIT algorithm can estimate the angles and Dopplers of two sources with very close DOAs or Dopplers but only when the other set of parameters are well separated. Additionally, the number of sources with the identical Doppler or DOA must be known or estimated in prior, which is difficult in practice. In the second general approach, a joint angle and frequency estimation (JAFE) method based on ESPRIT method has been proposed. The observed data vectors are smoothed spatially and temporally to construct a new high-dimensional data matrix and then the parameters can be estimated from the eigenvalues of the shifted matrix according to the rotation invariance principle [10]. The advantage of this method is that parameters of the same target are paired automatically and the computational burden is low compared to the MLE using a grid search. The disadvantage is that it is suboptimal and only formulated for a uniform linear array (ULA). Its performance achieves the CRLB at high SNR only when the smoothing factor is chosen to be some optimal value, which is never known in priori. Additionally, the computational complexity can still be quite high due to a high-dimensional eigen-decomposition [7], [8]. In [4] the JAFE algorithm has been extended to the colored noise, where it was pointed out that the root mean square error (RMSE) of the estimators based on subspace methods are far above the CRLB, even at high SNR [4].

In order to obtain higher estimation accuracy, the MLE has therefore been the subject of intense interest. In [5] two different 2-D approximations to the likelihood function were proposed to alleviate the effect of noise by exploiting a linear expansion of noise covariance matrix or by oblique projections

and a zero-forcing solution. The approximate 2-D MLE, as proposed, still requires a 2-D nonlinear optimization and hence, good initial estimates to avoid convergence to a local maximum are required.

In this paper we show how to compute the *exact MLE* in a computationally efficient manner. Previous work in this area by the authors [16], [17] have demonstrated good results for frequency estimation and also DOA estimation. These previous results are now extended to joint DOA/Doppler estimation. The key ingredient is the theorem of global optimization as discussed in [12]. Applying and extending this theorem to the problem at hand produces a method for global maximization of the compressed likelihood function. It replaces the multidimensional maximization by a multidimensional integration, which can be well approximated by Monte Carlo techniques [14], [15]. The Monte Carlo implementation uses importance sampling (IS), which has previously been shown to be a practical solution for these type of problems.

The paper is organized as follows. The data model and assumptions are described in Section II. In Section III, the likelihood function and the compressed likelihood function, which includes the parameters of interest, are derived. The global maximization technique is described in Section IV as well as the method to find the global maximum of the compressed likelihood function. Section V includes the choice of the importance function, the algorithm for DOAs and Dopplers, and discusses the effect of some algorithm parameters on the MSE of the IS estimator. Computer simulations and a comparison with other typical suboptimal methods are described in Section VI. Finally Section VII gives a discussion and conclusions about the proposed method.

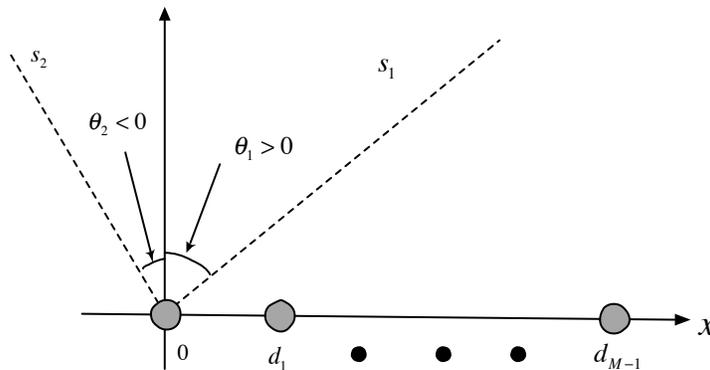


Fig. 1. Geometry of linear array for joint DOA and Doppler estimation.

II. DATA MODELS

Suppose there are p narrowband signal sources of interest at angles $\theta_1, \dots, \theta_p$ that are incident on a linear array of M sensors as shown in Figure 1. The possible range of arrival angles is $-\pi/2 \leq \theta_m \leq \pi/2, m = 1, \dots, p$ with $\theta_m = 0$ denoting the array normal direction. The leftmost sensor lies at the origin and the i th sensor position with respect to the origin is d_i . All signals have the same known carrier frequency and each signal has a different DOA and Doppler. The deterministic signals are given by complex signals as $s_i(n) = b_i e^{j2\pi f_i n}, i = 1, \dots, p$, where b_i is the deterministic complex amplitude and f_i is the Doppler frequency. After demodulation and downsampling, the signal received at the linear array, assuming a total of N snapshots is given by

$$\mathbf{x}(n) = \sum_{i=1}^p \mathbf{a}(\theta_i) s_i(n) + \mathbf{w}(n), \quad n = 0, 1, \dots, N-1, \quad (1)$$

where θ_i is the DOA of the i th signal and $\mathbf{a}(\theta_i)$ is the array manifold vector to a signal from direction θ_i and is defined as

$$\mathbf{a}(\theta_i) = \left[1 \quad e^{j2\pi f_c \frac{d_1}{c} \sin(\theta_i)} \quad \dots \quad e^{j2\pi f_c \frac{d_{M-1}}{c} \sin(\theta_i)} \right]^T \quad (2)$$

and f_c is the carrier frequency, and c is the speed of propagation. $\mathbf{w}(n)$ is the complex noise vector, which is assumed to be a sample vector of temporally and spatially white gaussian noise. As a result, we assume the known covariance matrix is $\sigma^2 \mathbf{I}$.

By suitable downsampling, the Doppler frequency takes values in the range $-0.5 \leq f_i < 0.5$. (1) can be written in matrix form as

$$\mathbf{x}(n) = \mathbf{A}(\boldsymbol{\theta}) \boldsymbol{\Phi}^n \mathbf{b} + \mathbf{w}(n), \quad n = 0, 1, \dots, N-1, \quad (3)$$

where $\mathbf{A}(\boldsymbol{\theta})$ and \mathbf{b} are defined as

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_p)], \quad (4)$$

$$\mathbf{b} = [b_1, b_2, \dots, b_p]^T, \quad (5)$$

$$\boldsymbol{\Phi} = \text{diag}\{\boldsymbol{\phi}\} = \text{diag}\{[e^{j\omega_1}, e^{j\omega_2}, \dots, e^{j\omega_p}]^T\}, \quad (6)$$

$$\boldsymbol{\phi} = [e^{j\omega_1}, e^{j\omega_2}, \dots, e^{j\omega_p}]^T, \quad (7)$$

and $\omega_i = 2\pi f_i$. This model can also be extended to the random signal model, where the complex amplitude b_i is varied randomly with time. Then the model (3) becomes

$$\mathbf{x}(n) = \mathbf{A}(\boldsymbol{\theta}) \boldsymbol{\Phi}^n \mathbf{b}(n) + \mathbf{w}(n), \quad n = 0, 1, \dots, N-1. \quad (8)$$

In this paper, only the deterministic model in (3) is considered with the random model to be studied in a future paper.

III. THE LIKELIHOOD FUNCTION

We define a vector $\boldsymbol{\eta}$ as the set of all unknown parameters, ie. $\eta_i = \{\theta_i, \omega_i, \text{Re}(b_i), \text{Im}(b_i)\}$, $i = 1, \dots, p$. The log likelihood function of the observed data $\mathbf{x}(n)$ is [5]

$$L(\mathbf{x}|\boldsymbol{\eta}) = -MN \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [\mathbf{x}(n) - \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}^n \mathbf{b}]^H [\mathbf{x}(n) - \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}^n \mathbf{b}], \quad (9)$$

and our goal is to maximize (9) over the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$. Unfortunately, $L(\mathbf{x}|\boldsymbol{\eta})$ is a very nonlinear function with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, but is quadratic with respect to the complex amplitude \mathbf{b} . As a result, we can maximize the likelihood function over \mathbf{b} by taking the gradient of $L(\mathbf{x}|\boldsymbol{\eta})$ with respect to \mathbf{b} and setting it to zero. When inserted back into the likelihood function, it will yield the *compressed likelihood function*. The complex derivative of the likelihood function with respect to \mathbf{b} (treating $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ as constants) can be written as

$$\begin{aligned} \frac{\partial L(\mathbf{x}|\boldsymbol{\eta})}{\partial \mathbf{b}^*} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [(\mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}^n)^H \mathbf{x}(n) - (\mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}^n)^H \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}^n \mathbf{b}] \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [(\boldsymbol{\Phi}^n)^H \mathbf{A}^H(\boldsymbol{\theta})\mathbf{x}(n) - (\boldsymbol{\Phi}^n)^H \mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})\boldsymbol{\Phi}^n \mathbf{b}]. \end{aligned} \quad (10)$$

According to Hadamard product property,

$$\mathbf{a}\mathbf{b}^T \odot \mathbf{Q} = \text{diag}(\mathbf{a})\mathbf{Q}\text{diag}(\mathbf{b}),$$

and $\boldsymbol{\Phi}$ is diagonal, then (10) can be simplified as

$$\begin{aligned} \frac{\partial L(\mathbf{x}|\boldsymbol{\eta})}{\partial \mathbf{b}^*} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [(\boldsymbol{\Phi}^n)^H \mathbf{A}^H(\boldsymbol{\theta})\mathbf{x}(n) - (\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \odot [\boldsymbol{\phi}^{*n}(\boldsymbol{\phi}^n)^T]) \mathbf{b}] \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [(\boldsymbol{\Phi}^n)^H \mathbf{A}^H(\boldsymbol{\theta})\mathbf{x}(n)] - \frac{N}{\sigma^2} (\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \odot \mathbf{E}(\boldsymbol{\omega})) \mathbf{b}, \end{aligned} \quad (11)$$

where the kl 'th element of the matrix \mathbf{E} is

$$\begin{aligned} [\mathbf{E}(\boldsymbol{\omega})]_{kl} &= \left[\frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{\phi}^{*n}(\boldsymbol{\phi}^n)^T \right]_{kl} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} e^{jn(\omega_l - \omega_k)} \\ &= \frac{1}{N} \mathbf{e}_k^H \mathbf{e}_l, \end{aligned} \quad (12)$$

$$\text{and } \mathbf{e}_k = \left[1 \quad e^{j\omega_k} \quad \dots \quad e^{j(N-1)\omega_k} \right]^T.$$

Let (11) equal zero so that the optimal estimate for \mathbf{b} is

$$\hat{\mathbf{b}} = [\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \odot \mathbf{E}(\boldsymbol{\omega})]^{-1} \mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega}), \quad (13)$$

where the vector $\mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega})$ is defined as

$$\mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega}) = \frac{1}{N} \sum_{n=0}^{N-1} [(\boldsymbol{\Phi}^n)^H \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{x}(n)]. \quad (14)$$

Furthermore $\mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega})$ can be shown to be decoupled, which means the i th elements of $\mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega})$ is only determined by the parameters of the i th source, that is,

$$\mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega}) = [r(\theta_1, \omega_1), \dots, r(\theta_p, \omega_p)]^T, \quad (15)$$

where

$$\begin{aligned} r(\theta, \omega) &= \frac{1}{N} \sum_{n=0}^{N-1} [e^{-j\omega n} \mathbf{a}^H(\theta) \mathbf{x}(n)] \\ &= \mathbf{a}^H(\theta) \frac{1}{N} \sum_{n=0}^{N-1} [e^{-j\omega n} \mathbf{x}(n)] \\ &= \mathbf{a}^H(\theta) \mathbf{x}(\omega), \end{aligned} \quad (16)$$

and $\mathbf{x}(\omega)$ is the normalized N-point discrete time Fourier transform of $\mathbf{x}(n)$ at frequency ω . $r(\theta, \omega)$ is the spatial-spectral transform of the observed data.

Substituting (13) into the likelihood function in (9) and omitting the constant terms or those terms not dependent on $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, produces the final compressed likelihood function as

$$L_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega}) = \frac{1}{\sigma^2} \mathbf{r}^H(\boldsymbol{\theta}, \boldsymbol{\omega}) [\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \odot \mathbf{E}(\boldsymbol{\omega})]^{-1} \mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega}). \quad (17)$$

Note that the number of sources is assumed to be known. However, this number can be estimated by one of the many model order selection techniques available [9], [19]

IV. GLOBAL MAXIMIZATION OF THE LIKELIHOOD FUNCTION

To implement the MLE we need to maximize (17) over $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$. Since the compressed likelihood function is nonlinear with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, a closed-form solution is not possible. A direct numerical maximization of (17) requires a $2p$ -D grid search, for which the computational burden increases exponentially with the number of sources. To avoid a direct maximization we propose to use the global maximization method by Pincus [12]. It provides a means of performing the nonlinear optimization and

is guaranteed to produce the global maximum. Based on the theorem given by Pincus, the \mathbf{x} that yields the global maximum of the cost function $g(\mathbf{x})$, is given by

$$\hat{x}_i = \lim_{\rho \rightarrow \infty} \frac{\int \dots \int x_i \exp(\rho g(\mathbf{x})) d\mathbf{x}}{\int \dots \int \exp(\rho g(\mathbf{y})) d\mathbf{y}} \quad i = 1, 2, \dots, M. \quad (18)$$

If we define

$$\bar{g}'(\mathbf{x}) = \frac{\exp(\rho g(\mathbf{x}))}{\int \dots \int \exp(\rho g(\mathbf{y})) d\mathbf{y}}$$

as the normalized function of $\exp(\rho g(\mathbf{x}))$, then $\bar{g}'(\mathbf{x})$ will satisfy $\int \dots \int \bar{g}'(\mathbf{x}) d\mathbf{x} = 1$ and can be considered as a pseudo-PDF of \mathbf{x} . So the optimal \hat{x}_i in (18) is the mean value given by

$$\hat{x}_i = \int \dots \int x_i \bar{g}'(\mathbf{x}) d\mathbf{x} \quad i = 1, \dots, M. \quad (19)$$

Intuitively, the pseudo-PDF $\bar{g}'(\mathbf{x})$ becomes a Dirac delta function at the location of the global maximum of $g(\mathbf{x})$ as $\rho \rightarrow \infty$.

The MLE for the parameters estimates, obtained from the location of the global maximum of the compressed likelihood function $L_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega})$ are given by

$$\hat{\theta}_i = \int \dots \int \theta_i \bar{L}'_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\theta} d\boldsymbol{\omega}, \quad (20)$$

$$\hat{\omega}_i = \int \dots \int \omega_i \bar{L}'_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\theta} d\boldsymbol{\omega}, \quad i = 1, \dots, M. \quad (21)$$

where $\bar{L}'_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega})$ is the normalized 2p-dimension pseudo-PDF of $\exp(\rho L_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega}))$,

$$\bar{L}'_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega}) = \frac{\exp(\rho L_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega}))}{\int \dots \int \exp(\rho L_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega})) d\boldsymbol{\theta} d\boldsymbol{\omega}}. \quad (22)$$

In practice, it is difficult to directly evaluate the multi-dimensional integral in (20) and (21), and it would seem that we have just traded one difficult problem for another. However, if we are able to generate in an easy manner T realizations of the vector $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, then (20) and (21) can be approximated by

$$\hat{\boldsymbol{\theta}} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\theta}(t), \quad (23)$$

$$\hat{\boldsymbol{\omega}} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\omega}(t) \quad (24)$$

where $\boldsymbol{\theta}(t)$ and $\boldsymbol{\omega}(t)$ are the t th realizations of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ distributed according to $\bar{L}'_c(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\omega})$. In this way the complicated integration can be approximated by simple sample averaging. Clearly, from the law of large numbers [20] the sample means will converge to the expected values as $T \rightarrow \infty$ so that there is no convergence issues. The only question in practice is the necessary value of T required for a

small error between the sample mean and the expected value. Note that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\omega}}$ are unbiased estimates of (20) and (21), and the variance will decrease as T increases.

It still remains to determine how realizations of the importance function can be easily generated. Since $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ is a highly nonlinear function of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, the direct generation of random number distributed according to $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ is difficult. To overcome this difficulty, as previously proposed [16], [17], one can use a simple PDF to generate realizations and then “unbias” the result using the concept of IS.

V. ON THE USE OF IMPORTANCE SAMPLING

To compute the integral of the function $h(\mathbf{x})$, where $\mathbf{x} \sim \bar{g}(\mathbf{x})$ is a PDF, one can use the relationship

$$\int h(\mathbf{x})\bar{g}(\mathbf{x})d\mathbf{x} = \int h(\mathbf{x})\frac{\bar{g}(\mathbf{x})}{\bar{p}(\mathbf{x})}\bar{p}(\mathbf{x})d\mathbf{x}, \quad (25)$$

where $\bar{p}(\mathbf{x})$ is another PDF, for which realizations of \mathbf{x} are more easily generated. We will call this new PDF the *the importance function*. If $\bar{p}(\mathbf{x})$ is chosen to be a simple function of \mathbf{x} so that realizations of \mathbf{x} can be easily generated, the integration in (25) can be approximated by the following averaging

$$\frac{1}{T} \sum_{t=1}^T h(\mathbf{x}_t) \frac{\bar{g}(\mathbf{x}_t)}{\bar{p}(\mathbf{x}_t)}, \quad (26)$$

where \mathbf{x}_t is the t th realization of \mathbf{x} and is distributed according to $\bar{p}(\mathbf{x})$. The performance of this approach is dependent upon the similarity of the shapes of the PDFs $\bar{p}(\mathbf{x})$ and $\bar{g}(\mathbf{x})$, and the value of T . Large T will of course increase the computational complexity, so small T is preferred. Also, to optimally “estimate” the integral, it can be shown that $\bar{p}(\mathbf{x})$ should be chosen similar to $\bar{g}(\mathbf{x})$. Another important point to keep in mind when choosing $\bar{p}(\mathbf{x})$ is that it should be simple enough so that $\mathbf{x} \sim \bar{p}(\mathbf{x})$ can be easily generated [13], [14]. Clearly, these are conflicting requirements so that some sort of compromise is required.

Assume that the importance function for the global optimization in (20) and (21) is $\bar{p}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$. The joint DOA and Doppler estimate is then given by

$$\hat{\theta}_i = \int \dots \int \theta_i \frac{\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})}{\bar{p}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})} \bar{p}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\theta} d\boldsymbol{\omega}, \quad (27)$$

$$\hat{\omega}_i = \int \dots \int \omega_i \frac{\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})}{\bar{p}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})} \bar{p}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\theta} d\boldsymbol{\omega}, \quad i = 1, \dots, M. \quad (28)$$

This method has been applied successfully to the frequency estimation of multiple sinusoidal signals in [16] and the DOA estimation of multiple stochastic sources in [17]. The choice of the importance function is discussed next.

A. Selection of the Importance Function

The compressed likelihood function $L_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ in (17) is nonlinear function of $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ due to the matrix inverse $[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \odot \mathbf{E}(\boldsymbol{\omega})]^{-1}$, so we must find an importance function $\bar{p}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ to generate realizations easily. A reasonable approach is to replace the inverse matrix in (17) with a diagonal matrix $\frac{1}{M}\mathbf{I}$, so the compressed likelihood function can be simplified to

$$\begin{aligned} L_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega}) &\approx \frac{1}{M\sigma_n^2} \mathbf{r}^H(\boldsymbol{\theta}, \boldsymbol{\omega}) \mathbf{r}(\boldsymbol{\theta}, \boldsymbol{\omega}) \\ &= \frac{1}{M\sigma_n^2} \sum_{i=1}^p |r(\theta_i, \omega_i)|^2. \end{aligned} \quad (29)$$

This type of approximation has been used in [16] and [17] and in fact, yields the true compressed likelihood function for well separated sources. Although the well separated case is not the problem of interest here, it can occur for some combinations of DOA/Dopplers of multiple sources. An importance function for (20) and (21) in this case is

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\omega}) &= \exp\left(\rho_1 \frac{1}{M\sigma_n^2} \sum_{i=1}^p |r(\theta_i, \omega_i)|^2\right) \\ &= \prod_{i=1}^p \exp(\rho_1 I_x(\theta_i, \omega_i)). \end{aligned} \quad (30)$$

where

$$I_x(\theta_i, \omega_i) = \frac{1}{M\sigma_n^2} |r(\theta_i, \omega_i)|^2, \quad (31)$$

and ρ_1 is another constant different from ρ , which we will discuss later. Note that $I_x(\theta, \omega)$ is the two-dimensional Fourier transform of the array snapshots $\mathbf{x}(n)$, $n = 0, 1, \dots, N - 1$. We normalize the importance sampling function $p(\boldsymbol{\theta}, \boldsymbol{\omega})$ by

$$\bar{p}(\boldsymbol{\theta}, \boldsymbol{\omega}) = \frac{\prod_{i=1}^p \exp(\rho_1 I_x(\theta_i, \omega_i))}{\int \dots \int \prod_{i=1}^p \exp(\rho_1 I_x(\theta_i, \omega_i)) d\theta_i d\omega_i} \quad (32)$$

to obtain the normalized importance function.

There are two advantages for choosing the importance function as (30). One is that the importance function is decoupled, that is, the original $2p$ dimensional PDF $\bar{p}(\boldsymbol{\theta}, \boldsymbol{\omega})$ is reduced to the products of p identical 2-dimensional PDFs. It is then much easier to generate realizations of the parameters. The second advantage is that the simplified function in (29) is a good approximation to the compressed likelihood function in (17) when all or even some of θ_i or ω_i are well separated, i.e., the distance

between the DOAs and/or Dopplers is larger than the Rayleigh resolution limits. Why this is the case is described next for the ULA.

The kl th element of matrix $\mathbf{A}(\boldsymbol{\theta})^H \mathbf{A}(\boldsymbol{\theta})$ is given by

$$\begin{aligned} [\mathbf{A}(\boldsymbol{\theta})^H \mathbf{A}(\boldsymbol{\theta})]_{kl} &= \mathbf{a}(\theta_k)^H \mathbf{a}(\theta_l) \\ &= \sum_{i=1}^M e^{j2\pi f_c (i-1)d/c(\sin(\theta_l) - \sin(\theta_k))}. \end{aligned} \quad (33)$$

If the element spacing of ULA is chosen as half-wavelength at the carrier frequency, i.e., $f_c d/c = 0.5$, we can prove that

$$|[\mathbf{A}(\boldsymbol{\theta})^H \mathbf{A}(\boldsymbol{\theta})]_{kl}| \leq 1, \quad \text{if } |\sin(\theta_l) - \sin(\theta_k)| > \frac{1}{M}, k \neq l, \quad (34)$$

and then $\mathbf{A}(\boldsymbol{\theta})^H \mathbf{A}(\boldsymbol{\theta}) \approx M\mathbf{I}$. In a similar way, we can prove

$$|[\mathbf{E}(\boldsymbol{\omega})]_{kl}| \leq \frac{1}{N}, \quad k \neq l, \quad (35)$$

if the Doppler spacing is much greater than $1/N$ and $\mathbf{E}(\boldsymbol{\omega}) \approx \mathbf{I}$. Therefore, only if the number of elements or snapshots is large enough, is the simplified compressed likelihood function in (29) equal to the true compressed likelihood function.

B. Parameter estimation by importance sampling

Based on the importance function defined in (32), the estimators of DOA and Doppler using IS in (27) and (28) are

$$\hat{\theta}_i = \frac{1}{T} \sum_{t=1}^T [\boldsymbol{\theta}_t]_i \frac{\bar{L}_c(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)}{\bar{p}(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)}, \quad (36)$$

$$\hat{\omega}_i = \frac{1}{T} \sum_{t=1}^T [\boldsymbol{\omega}_t]_i \frac{\bar{L}_c(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)}{\bar{p}(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)}, \quad (37)$$

where $\boldsymbol{\theta}_t$ and $\boldsymbol{\omega}_t$ are the t th realizations of the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ distributed according to the PDF $\bar{p}(\boldsymbol{\theta}, \boldsymbol{\omega})$ in (32). It is noted that since the DOAs and Dopplers are bounded from below and above, it is more convenient to evaluate the mean using the circular mean, rather than the linear mean. It has been shown that the linear mean will result in a biased estimator, especially at low SNR and/or short data lengths [18]. Additionally, the use of the circular mean will reduce the computation greatly as will be explained shortly.

To describe the concept of a circular mean first consider a circular random variable. This is one that takes values on a finite interval and that finite interval can be considered to be mapped into the

unit circle. Specifically, if a random variable ϕ is defined on the interval $[0, 1]$ with the PDF $p(\phi)$, the circular mean for ϕ is defined as

$$E_c[\phi] = \frac{1}{2\pi} \angle \int_0^1 \exp(j2\pi\phi) p(\phi) d\phi, \quad (38)$$

where \angle denotes the angle in radians. If we have T realizations of ϕ , the sample circular mean for ϕ is

$$\hat{\phi}_c = \frac{1}{2\pi} \angle \frac{1}{T} \sum_{t=1}^T \exp(j2\pi\phi_t). \quad (39)$$

If the DOAs and Dopplers are not within the interval $[0, 1]$, they may be easily mapped into this interval. The resulting transformed parameters can be estimated and then inverse mapped to obtain estimates of the original parameters. This is valid due to the invariance property of the MLE. If we define $\bar{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\omega}}$ as the normalized DOAs and Dopplers within $[0, 1]$, the MLE of DOA and Doppler with the circular means in (36) and (37) are

$$\bar{\theta}_{i,mle} = \frac{1}{2\pi} \angle \frac{1}{T} \sum_{t=1}^T \frac{\bar{L}_c(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)}{\bar{p}(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)} \exp(j2\pi[\bar{\boldsymbol{\theta}}_t]_i), \quad (40)$$

$$\bar{\omega}_{i,mle} = \frac{1}{2\pi} \angle \frac{1}{T} \sum_{t=1}^T \frac{\bar{L}_c(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)}{\bar{p}(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)} \exp(j2\pi[\bar{\boldsymbol{\omega}}_t]_i), \quad (41)$$

where $\bar{\boldsymbol{\theta}}_t$ and $\bar{\boldsymbol{\omega}}_t$ are the t th realizations of vector $\bar{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\omega}}$ distributed according to the PDF $\bar{p}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\omega}})$ in (32).

There are some tricks that can be employed to make the computation more manageable. The first is used to reduce the computation greatly. It is noted that we need only determine the angle of the complex quantity in (39), and therefore the normalization constant term in computing $\bar{L}_c(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)$ and $\bar{p}(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)$ can be omitted in (40) and (41). Another trick is necessary to reduce any computational overflow associated with evaluation of exponential forms. In particular, the exponential forms in $L_c(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)$ and $p(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)$ can lead to overflow. To avoid this we replace the original estimates by the normalized versions as

$$\bar{\theta}_{i,mle} = \frac{1}{2\pi} \angle \frac{1}{T} \sum_{t=1}^T w'(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) \exp(j2\pi[\bar{\boldsymbol{\theta}}_t]_i), \quad (42)$$

$$\bar{\omega}_{i,mle} = \frac{1}{2\pi} \angle \frac{1}{T} \sum_{t=1}^T w'(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) \exp(j2\pi[\bar{\boldsymbol{\omega}}_t]_i), \quad (43)$$

where the weight coefficient is defined as

$$w'(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) = \exp \left[\rho L_c(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \sum_{i=1}^P \rho_1 I_x([\bar{\boldsymbol{\theta}}_t]_i, [\bar{\boldsymbol{\omega}}_t]_i) - \max_{1 \leq k \leq T} \left(\rho L_c(\bar{\boldsymbol{\theta}}_k, \bar{\boldsymbol{\omega}}_k) - \sum_{i=1}^P \rho_1 I_x([\bar{\boldsymbol{\theta}}_k]_i, [\bar{\boldsymbol{\omega}}_k]_i) \right) \right] \quad (44)$$

and $I_x(\theta_i, \omega_i)$ is defined in (31). Since we need only take the angle of a complex quantity, any scaling will not affect the result.

C. Summary of steps

For sake of clarity the entire procedure of joint estimation of both DOA and Doppler by importance sampling is summarized by the following steps:

- 1) Based on the observed data $\mathbf{x}(n)$, $n = 0, \dots, N-1$, evaluate the 2-D spatial-temporal periodogram $I(\theta, \omega)$ according to (16) and (31) to obtain the discretized importance function

$$\bar{p}(\bar{\theta}_l, \bar{\omega}_k) = \frac{\exp(\rho_1 I(\bar{\theta}_l, \bar{\omega}_k))}{\sum_{l=1}^L \sum_{k=1}^K \exp(\rho_1 I(\bar{\theta}_l, \bar{\omega}_k))}, \quad -1/2 \leq \bar{\theta}_l, \bar{\omega}_k \leq 1/2, \quad (45)$$

where L and K are the total number of points for DOA and Doppler, respectively. Typically, 20 times the number of elements and snapshots is sufficient, that is, $L = 20M$ and $K = 20N$.

- 2) Generate a realization of parameter vector pair $(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)$ using *inverse probability integral transformation* [20]. Note that it is important that at least one parameter in the different parameter pairs $([\boldsymbol{\theta}_t]_i, [\boldsymbol{\omega}_t]_i)$ is different from the others to ensure that the matrix $\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \odot \mathbf{E}(\boldsymbol{\omega})$ is invertible.
- 3) Repeat step 2) to obtain T realizations of the vector pair $(\boldsymbol{\theta}_t, \boldsymbol{\omega}_t)$.
- 4) Evaluate the weight coefficient $w'(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)$ defined in (44) and estimate DOAs and Dopplers using (42) and (43), where the compressed likelihood function is given by (17).
- 5) Convert the normalized parameter pairs $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\omega}})$ into the true parameter pairs $(\boldsymbol{\theta}, \boldsymbol{\omega})$.

One straightforward way to generate random numbers in step 2) is based on the inverse probability transformation as applied to 2-D [20]. In this approach one generates two independent identically distributed uniform random numbers, then search for the point from the 2-D cumulative function of $\bar{p}(\bar{\theta}_l, \bar{\omega}_k)$ to match the generated uniform random numbers. It requires a 2-D search to find the optimal match and is computationally prohibitive. A simpler method is to decompose the 2-D PDF into the product of two 1-D PDFs. It is easily shown that $\bar{p}(\bar{\theta}_l, \bar{\omega}_k)$ can be described as

$$\bar{p}(\bar{\theta}_l, \bar{\omega}_k) = \bar{p}(\bar{\theta}_l | \bar{\omega}_k) \bar{p}(\bar{\omega}_k) = \bar{p}(\bar{\omega}_k | \bar{\theta}_l) \bar{p}(\bar{\theta}_l), \quad (46)$$

so the 2-D PDF can be considered as the product of two 1-D PDFs. One feasible way is to evaluate the cumulative function $c(\bar{\omega})$ and $c(\bar{\theta} | \bar{\omega})$ from the 2-D PDF $\bar{p}(\bar{\theta}, \bar{\omega})$ is to generate one uniform random number ϕ_1 in $[0, 1]$ and find the parameter $\bar{\omega}_1$ to make $c(\bar{\omega}_1)$ the closest to ϕ_1 . Next, generate another uniform random number ϕ_2 and find the optimal parameter $\bar{\theta}_1$ to make $c(\bar{\theta}_1 | \bar{\omega}_1)$ closest to ϕ_2 . Continue generating $(\bar{\theta}_2, \bar{\omega}_2), \dots, (\bar{\theta}_p, \bar{\omega}_p)$ in a similar way with the condition that all $\bar{\omega}_i$ s are distinct. This simplification reduces the computational burden at the cost of generality. Two sources with identical Doppler but different DOAs cannot be identified, but this case rarely occurs.

D. Performance analysis of the IS estimators

For the integration in (25), it is easily proven that the IS estimator in (26) is unbiased, and the variance of the estimator is finite when the expectation satisfies

$$E_{\bar{p}} \left[h^2(\mathbf{x}) \frac{\bar{g}'^2(\mathbf{x})}{\bar{p}^2(\mathbf{x})} \right] = \int h^2(\mathbf{x}) \frac{\bar{g}'^2(\mathbf{x})}{\bar{p}(\mathbf{x})} d\mathbf{x} < \infty, \quad (47)$$

where $E_{\bar{p}}$ is the expectation operation with respect to PDF $\bar{p}(\mathbf{x})$ [15]. If the choice of the importance function $\bar{p}(\mathbf{x})$ is

$$\bar{p}(\mathbf{x}) = \frac{|h(\mathbf{x})|\bar{g}'(\mathbf{x})}{\int |h(\mathbf{z})|\bar{g}'(\mathbf{z})d\mathbf{z}}, \quad (48)$$

then it can be shown that the variance of estimator is minimum [13]. However, (48) is unrealizable because the optimal importance function is just the original function which we sought to avoid considering as a PDF. However, it indicates how to choose the importance function. One rule is to find the distribution function \bar{p} to keep $h\bar{g}'/\bar{p}$ flat for all possible parameter values, or try to keep $h\bar{g}'/\bar{p}$ finite if the previous condition cannot be held. Another consideration, which should also be kept in mind, is that random numbers with the distribution function to be the importance function are generated easily.

In order to easily generate a realization, the importance function for the cost function $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ in (22) is chosen as $\bar{p}'(\boldsymbol{\theta}, \boldsymbol{\omega})$ in (32). There are three parameters that need to be carefully chosen — ρ , ρ_1 , and T in the IS method. As ρ goes to infinity, $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ approximates a Dirac delta function at the true value, but the parameter ρ cannot be chosen too large or computational overflow will occur. The role of ρ_1 is to amplify $I(\boldsymbol{\theta}, \boldsymbol{\omega})$ in (32) so that $\bar{p}'(\boldsymbol{\theta}, \boldsymbol{\omega})$ has a sharp peak at the maximum of $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$. The number of realizations, T , is dependent on the choice of the importance function. A good choice of the importance function will lead to small T and thus, less computation. As discussed previously, $\bar{p}'(\boldsymbol{\theta}, \boldsymbol{\omega})$ is an approximation to $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ and their peak locations will be different if parameter pairs of two sources are within the Rayleigh resolution limits. A large ρ_1 will decrease the probability where random numbers include the true value, and thus a large number of realizations T is required to keep the same variance — hence, increasing the computational burden. If all parameters are well separated, $\bar{p}'(\boldsymbol{\theta}, \boldsymbol{\omega})$ is a good approximation to $\bar{L}'_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$, and as discussed in subsection A has a sharp peak at the true parameter locations. If there are two sources with almost the same DOA and Doppler, the situation becomes more complicated. First, the peak of the compressed likelihood function $L_c(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\omega})$ is not as evident as before so that a larger ρ is required. Secondly, the importance function is a biased approximation, so that a small value of ρ_1 can increase the probability of inclusion of true parameters.

VI. COMPUTER SIMULATIONS

In this section, two proposed DOA/Doppler estimation algorithms, JAFE [8] and FSF-ESPRIT [6], are compared with the proposed IS method. We choose an 8-element ULA with half-wavelength spacing since JAFE and FSF-ESPRIT are only suitable for ULA. Note, however, that our method can easily be extended to a nonuniform linear array and other known array shapes. For this case, the 3 dB beamwidth of the beamformer with uniform shading weights is 12.8° and the Doppler Rayleigh resolution limit is $2/N$, where N is the number of snapshots. For JAFE, the temporal and spatial smoothing factors are chosen as 5 and 2, respectively. The number of sources with the same Doppler or DOA is unknown in practice, so all sources are assumed to have different parameters for FSF-ESPRIT. The mean square error (MSE) of the three estimators are compared with the Cramer-Rao lower bound (CRLB). See the Appendix for a derivation of the CRLB.

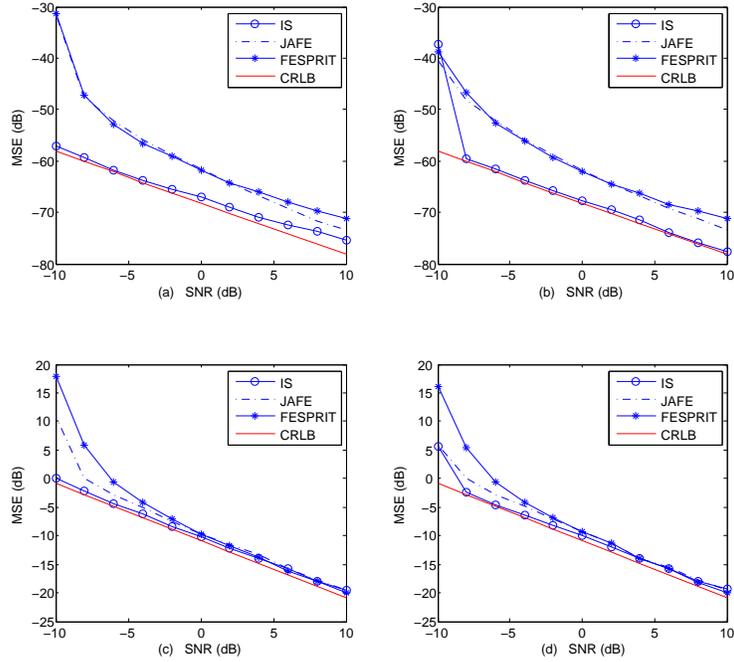


Fig. 2. The MSE of Doppler ((a) and (b)) and DOA ((c) and (d)) estimates for three algorithms vs SNR when sources are resolvable with $(-8^\circ, 0.15)$ and $(10^\circ, 0.207)$. Results are obtained by averaging of 1000 Monte Carlo simulations with $N = 50$, $\rho = 80$ and $\rho_1 = 6$.

Example 1. We address the performance of all algorithms versus SNR and the number of snapshots when the sources are resolvable. First, the two sources with the parameters $(-8^\circ, 0.15)$ and $(10^\circ, 0.207)$ are chosen, and $N = 50$ snapshots, $\rho = 80$, $\rho_1 = 6$ are used. The MSE for the three estimators are

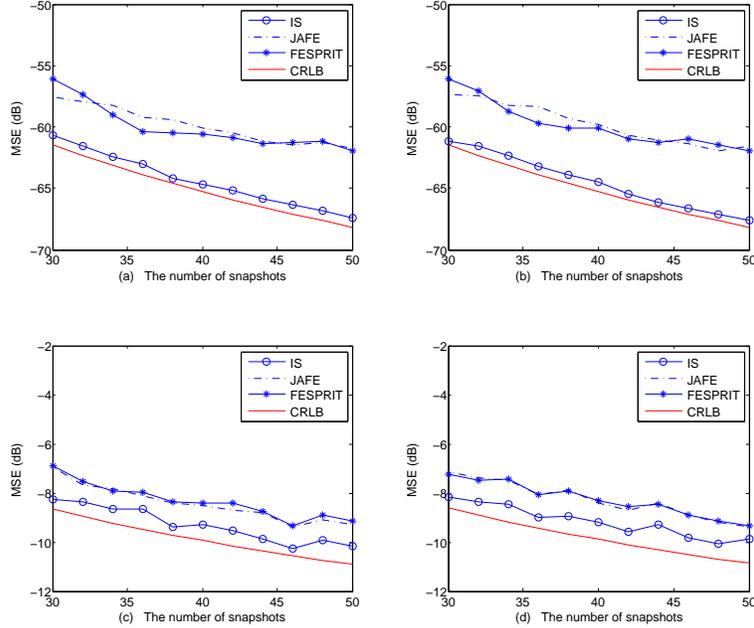


Fig. 3. The MSE of Doppler ((a) and (b)) and DOA ((c) and (d)) estimates for three algorithms vs N when sources are resolvable with $(-8^\circ, 0.15)$ and $(10^\circ, 0.207)$. Results are obtained by averaging of 1000 Monte Carlo simulations with $\text{SNR} = 0\text{dB}$, $\rho = 80$ and $\rho_1 = 6$.

determined from the average of 1000 Monte Carlo simulations, and the results are shown in Figure 2 as SNR changes from -10 dB to 10 dB. The realization number T is chosen as 2000. It can be seen from (a) and (b) in figure 2 that the estimation accuracy of Doppler by IS is very close to the CRLB for almost all SNRs. The JAFE and FSF-ESPRIT methods, on the other hand, perform poorly for Doppler estimation (see Figure 2a and 2b) and even for DOA estimation (see Figures 2c and 2d) have SNR thresholds well above that for IS. In all cases the IS approach produces more accurate estimates for DOA and Doppler. The performance of the JAFE and FSF-ESPRIT estimators agree with those presented in [6].

Next, we fix $\text{SNR} = 0$ dB and keep other conditions the same as before. In Figure 3 the MSE of the three estimators are plotted versus the number of snapshots. Again it is quite apparent that IS has better performance than the competing techniques, and furthermore its performance is close to CRLB even for a small number of snapshots.

Example 2. Here the performance of three estimators are compared for two sources with closely spaced parameters. The parameter pairs of the two sources are $(-8^\circ, 0.15)$ and $(-5^\circ, 0.158)$ and $N = 50$. The two sources are within the Rayleigh resolution limits for DOA and Doppler, so the importance function in (32) is a poor approximation to the true function given in (22). A large value of ρ , a small

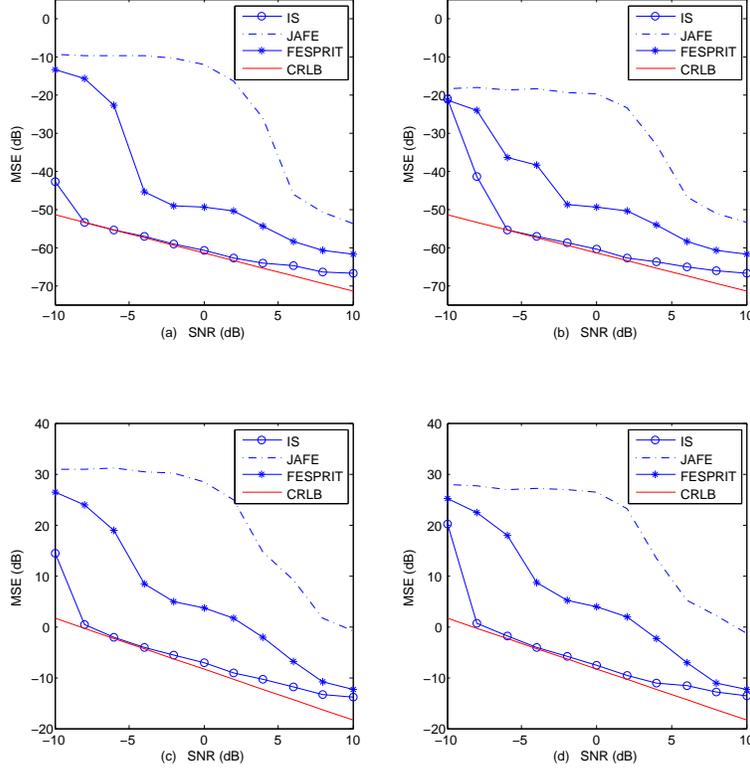


Fig. 4. The MSE of Doppler ((a) and (b)) and DOA ((c) and (d)) estimates for three algorithms vs SNR when sources are close with $(-8^\circ, 0.15)$ and $(-5^\circ, 0.158)$. Results are obtained by averaging of 1000 Monte Carlo simulations with $N = 50$, $\rho = 400$ and $\rho_1 = 4$.

value of ρ_1 , and a large value T is therefore necessary for the IS method to produce good results with a reasonable number of realizations. In Figure 4 we plot the dependence of MSE for the three algorithms vs SNR, where the number of realizations is $T = 8000$ and $\rho_1 = 4$, $\rho = 400$. In this case, JAFE and FSF-ESPRIT have a large bias for DOA and Doppler estimation due to many outliers which occur at the low SNR of 0 dB that was chosen, but the MSE of the IS estimator is still very close to the CRLB. For an SNR above 5 dB, the performance of the IS is slightly higher than the CRLB. The reason is that the approximation of the likelihood function by the chosen importance function is poorer as the SNR increases. To compensate for this tendency a smaller ρ_1 and larger T are required but in this experiment we have fixed ρ_1 and T .

Example 3. As a further demonstration of the good performance of the IS approach, we next consider three sources. First, we will examine the performance of all algorithms when only one parameter of DOA and Doppler is close. For three sources whose true parameter pairs are listed in Table I, Dopplers of source 1 and 2 are very close but their DOAs are separated, and the DOAs for source 2 and 3 are

TABLE I
MSE RELATIVE TO CRLB OF THREE ALGORITHMS FOR 3 SOURCES WITH ONLY ONE CLOSE PARAMETER

Source	DOA			Doppler		
	1	2	3	1	2	3
True parameters	-8°	10°	12°	0.150	0.158	0.240
Relative MSE of JAFE (dB)	18.45	18.66	2.0	12.3	20.2	14.6
Relative MSE of FSF (dB)	28.6	17.84	2.0	22.0	22.9	12.9
Relative MSE of IS (dB)	5.54	5.75	6.0	5.8	6.2	6.0

close but their Dopplers are quite different. We fix the algorithm parameters as $\text{SNR} = 0$ dB, $N = 50$, $\rho = 200$ and $\rho_1 = 6$. The MSE of the three algorithms *relative to* the CRLB are listed in Table I, where the increase in dB of the MSE over the CRLB is shown. It can be seen that JAFE and FSF-ESPRIT estimators cannot estimate well for two sources with close Doppler. However, the IS method has almost the same MSE for all sources.

TABLE II
MSE RELATIVE TO CRLB OF THREE ALGORITHMS FOR 4 SOURCES WITH TWO CLOSE PARAMETER

Source	DOA				Doppler			
	1	2	3	4	1	2	3	4
True parameters	-8°	-5°	12°	14°	0.150	0.158	-0.240	-0.230
Relative MSE of JAFE (dB)	39.8	39.1	40.6	40.6	58.6	58.2	57.7	57.8
Relative MSE of FSF (dB)	39.4	38.7	39.6	40.2	58.4	58.3	57.1	57.6
Relative MSE of IS (dB)	13.5	13.7	14.6	13.1	13.3	14.7	13.6	14.4

We add source 4 and change parameters of the other sources as listed in Table II. There are two groups and each group has two close sources. The parameter for IS method is chosen as $\rho = 400$, $\rho_1 = 5$, $T = 8000$, and $\text{SNR} = 5$ dB. The number of snapshots is 50 and the MSE relative to CRLB for three algorithms are listed in Table II in dB. It can be seen that only IS method can identify four sources in this case although the deviation of MSE to CRLB is a little higher than before, but JAFE and FSF have large MSE due to large outliers using 1000 estimates.

VII. DISCUSSION AND CONCLUSIONS

We have proposed a computationally modest technique to jointly estimate DOAs and Dopplers of narrowband sources using an MLE. It is noticed that our method does not require initial parameter

estimates but can still guarantee that the global maximum of the likelihood function will be found and hence the exact MLE. The computation is far less relative to a grid search that requires an exponentially increasing number of evaluations of the likelihood function for multiple sources. Relative to other proposed methods such as JAFE and FSF-ESPRIT, the IS approach exhibits better performance at low SNR and for a small number of snapshots. For closely spaced sources, for which all parameters are within the Rayleigh resolution limits, the IS algorithm is able to produce reliable estimates but requires an increasing number of realizations. For this important case the other methods fail.

APPENDIX

Use the original likelihood function in (9) as the general model, then the Fisher Information Matrix (FIM) can be computed simply according to theory in [11]. Since \mathbf{b} is complex, we extend the parameter number to $4p$ by including the real part and imaginary part of \mathbf{b} . Define the new parameter vector $\boldsymbol{\eta}$ as $\boldsymbol{\eta} = [\boldsymbol{\theta}^T, \boldsymbol{\omega}^T, \text{Re}(\mathbf{b}^T), \text{Im}(\mathbf{b}^T)]$ and the gradient vector of $\mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}$ with respect to $\boldsymbol{\eta}$ is defined as

$$\mathbf{D}_n(\boldsymbol{\eta}) = \left[\frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\eta_1}, \frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\eta_2}, \dots, \frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\eta_{4p}} \right], \quad (49)$$

then the final FIM is given by

$$\mathbf{I}(\boldsymbol{\eta}) = \frac{2}{\sigma_n^2} \text{Re} \left(\sum_{n=0}^{N-1} \mathbf{D}_n^H(\boldsymbol{\eta}) \mathbf{D}_n(\boldsymbol{\eta}) \right). \quad (50)$$

The CRLB for the i th parameter η_i is obtained from the inverse of the FIM as

$$\text{CRLB}(\eta_i) = [\mathbf{I}^{-1}(\boldsymbol{\eta})]_{ii}. \quad (51)$$

Define

$$\begin{aligned} \mathbf{D}_\theta &= \frac{\partial \mathbf{a}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \left[\frac{\partial \mathbf{a}(\theta_1)}{\partial \theta_1}, \frac{\partial \mathbf{a}(\theta_2)}{\partial \theta_2}, \dots, \frac{\partial \mathbf{a}(\theta_p)}{\partial \theta_p} \right], \\ \mathbf{B} &= \text{diag}(\mathbf{b}). \end{aligned}$$

Then the derivative of $\mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}$ with respect to each parameter can be evaluated as follows

$$\begin{aligned} \frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\partial \boldsymbol{\theta}^T} &= \mathbf{D}_\theta \Phi^n \mathbf{B}, \\ \frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\partial \boldsymbol{\omega}^T} &= jn \mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B}, \\ \frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\partial \text{Re}(\mathbf{b}^T)} &= \mathbf{A}(\boldsymbol{\theta}) \Phi^n, \\ \frac{\partial \mathbf{A}(\boldsymbol{\theta})\Phi^n\mathbf{b}}{\partial \text{Im}(\mathbf{b}^T)} &= j \mathbf{A}(\boldsymbol{\theta}) \Phi^n. \end{aligned}$$

So the $\mathbf{D}_n^H(\boldsymbol{\eta})\mathbf{D}_n(\boldsymbol{\eta})$ can be expressed by

$$\mathbf{D}_n^H(\boldsymbol{\eta})\mathbf{D}_n(\boldsymbol{\eta}) = \begin{bmatrix} (\mathbf{D}_\theta \Phi^n \mathbf{B})^H \\ (jn \mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B})^H \\ (\mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H \\ (j \mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H \end{bmatrix} \begin{bmatrix} (\mathbf{D}_\theta \Phi^n \mathbf{B})^H \\ (jn \mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B})^H \\ (\mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H \\ (j \mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H \end{bmatrix}^H \quad (52)$$

and the FIM is

$$\mathbf{I}(\boldsymbol{\eta}) = \frac{2}{\sigma_n^2} \text{Re} \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{21}^H & \mathbf{I}_{31}^H & j\mathbf{I}_{31}^H \\ \mathbf{I}_{21} & \mathbf{I}_{22} & -j\mathbf{I}_{32}^H & \mathbf{I}_{32}^H \\ \mathbf{I}_{31} & j\mathbf{I}_{32} & \mathbf{I}_{33} & j\mathbf{I}_{33}^H \\ -j\mathbf{I}_{31} & \mathbf{I}_{32} & -j\mathbf{I}_{33} & \mathbf{I}_{33} \end{bmatrix} \quad (53)$$

where

$$\begin{aligned} \mathbf{I}_{11} &= \sum_{n=0}^{N-1} (\mathbf{D}_\theta \Phi^n \mathbf{B})^H \mathbf{D}_\theta \Phi^n \mathbf{B} = (\mathbf{B}^H \mathbf{D}_\theta^H \mathbf{D}_\theta \mathbf{B}) \odot N\mathbf{E} \\ \mathbf{I}_{22} &= \sum_{n=0}^{N-1} n^2 (\mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B})^H \mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B} = (\mathbf{B}^H \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}) \mathbf{B}) \odot \sum_{n=0}^{N-1} n^2 \phi^{n*} (\phi^n)^T \\ \mathbf{I}_{33} &= \sum_{n=0}^{N-1} (\mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H \mathbf{A}(\boldsymbol{\theta}) \Phi^n = \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}) \odot N\mathbf{E} \\ \mathbf{I}_{21} &= \sum_{n=0}^{N-1} (jn \mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B})^H \mathbf{D}_\theta \Phi^n \mathbf{B} = -j \mathbf{B}^H \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{D}_\theta \mathbf{B} \odot \sum_{n=0}^{N-1} n \phi^{n*} (\phi^n)^T \\ \mathbf{I}_{31} &= \sum_{n=0}^{N-1} (\mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H \mathbf{D}_\theta \Phi^n \mathbf{B} = \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{D}_\theta \mathbf{B} \odot N\mathbf{E} \\ \mathbf{I}_{32} &= \sum_{n=0}^{N-1} (\mathbf{A}(\boldsymbol{\theta}) \Phi^n)^H n \mathbf{A}(\boldsymbol{\theta}) \Phi^n \mathbf{B} = \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}) \mathbf{B} \odot \sum_{n=0}^{N-1} n \phi^{n*} (\phi^n)^T. \end{aligned}$$

These results allow the computation of the FIM so that we finally obtain the CRLB of (51).

REFERENCES

- [1] H. Krim and M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Mag.*, pp.67-94, July 1996.
- [2] S. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, N.J., 1988.
- [3] F. Athley, "Asymptotically Decoupled Angle-Frequency Estimation with Sensor Arrays", *Pro. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May, 1999, pp. 1098–1102.
- [4] M. Viberg and P. Stoica, "A Computationally Efficient Method for Joint Direction Finding and Frequency Estimation in Colored Noise", *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May, 1998, pp. 1547–1551.
- [5] M. Djeddou, A. Belouchrani, and A. Aouada, "Maximum Likelihood Angle-Frequency Estimation in Partially Known Correlated Noise for Low-Elevation Targets", *IEEE Trans. Signal Processing*, vol.53, pp.3057-3064, August 2005.
- [6] J.D. Lin, W.H. Fang, et. al, "FSF MUSIC for Joint DOA and Frequency Estimation and Its Performance Analysis", *IEEE Trans. Signal Processing*, vol.54, pp.4529-4542, Dec. 2006.
- [7] A. Lemma, A. V. D. Veen, and E. F. Deprettere, "Joint Angle-Frequency Estimation Using Multi-resolution ESPRIT," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May, 1998, pp. 1957–19601.
- [8] A. Lemma, A. V. D. Veen, and E. Deprettere, "Analysis of Joint Angle-Frequency Estimation Using ESPRIT," *IEEE Trans. Signal Processing*, vol. 51, pp. 1264-1283, May 2003.
- [9] M. Wax and T. Kailath, "Detection of Signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp.387-392, Apr. 1985.
- [10] R. Roy and T. Kailath, "ESPRIT- Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp.984-995, July 1989.
- [11] S. Kay, *Fundamentals of Statistical Signal Processing , Estimation Theory* , Prentice Hall, Englewood Cliffs, NJ, 1993.
- [12] M. Pincus, "A Closed Form Solution for Certain Programming Problems", *Oper. Research*, pp. 690-694, 1962.
- [13] R. Rubinstein, *Monte Carlo optimization, simulation, and sensitivity of queuing networks*, Wiley series in probability and mathematical statistics, 1986.
- [14] M. Kalos, P. Whitlock, *Monte Carlo Methods*, J. Wiley and Sons, 1986.
- [15] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd edition, pg. 90, Springer, 2004
- [16] S. Kay and S. Saha, "Mean Likelihood Frequency Estimation", *IEEE Trans. Signal Processing*, vol.48, pp.1937-1946, July 2000.
- [17] S. Saha, S. Kay, and H. Wang, "An Exact Maximum Likelihood Narrowband Direction of Arrival Estimator", Submitted to *IEEE Trans. Signal Processing*.
- [18] S. Kay, "Comments on Frequency Estimation by Linear Prediction", *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp.198-199, April 1979.
- [19] C. Xu, S. Kay, M. Rangaswamy, F. Lin, "Model Order Estimation for Adaptive Radar Clutter Cancellation", *IEEE Int. Conference on Waveform and Diversity*, Pisa, Italy, 2007.
- [20] S. Kay, *Intuitive Probability and Random Processes using MATLAB*, Springer, New York, 2005.