Particle Filters for Tracking an Unknown Number of Sources
EDICS: 1-DETC

Jean-René Larocque, James P. Reilly* and William Ng

Department of Electrical and Computer Engineering,

McMaster University, 1280 Main St. W.,

Hamilton, Ontario,

Canada L8S 4K1

Submitted to IEEE Transactions on Signal Processing, May, 2001.

Abstract

This paper addresses the application of sequential importance sampling (SIS) schemes to tracking DOAs of an unknown number of sources, using a passive array of sensors. This proposed technique has significant advantages in this application, including the ability to detect a changing number of signals at arbitrary times throughout the observation period, and that the requirement for quasi-stationarity over a limited interval may be relaxed.

We propose the use of a reversible jump MCMC [1] step to enhance the statistical diversity of the particles. This step also enables us to introduce two novel moves which significantly enhance the performance of the algorithm when the DOA tracks cross. The superior performance of the method is demonstrated by examples of application of the particle filter to sequential tracking of the DOAs of an unknown and non-stationary number of sources, and to a scenario where the targets cross. Our results are compared to the PASTd method [2].

*Permission to publish abstract separately is granted.

J. Reilly, corresponding author: ph: 905 525 9140 x22895, fax: 905 521 2922, email: reillyj@mcmaster.ca

I. Introduction

The problem of tracking the directions of arrival (DOAs) of multiple targets in background noise using passive arrays of sensors is of great interest to the signal processing community, with applications in communications, radar, sonar, acoustics and others. For example, in a beamforming application, one is typically interested in extracting a signal of interest arriving onto an array of sensors, from multiple interfering sources arriving from different DOAs. For this approach to be effective, the DOA of the desired source must be estimated from the received data. In many scenarios, the desired source is moving, necessitating *target tracking* of the desired DOA.

Recently, many DOA estimation techniques have been proposed. These include beamforming methods [3][4], subspace-based methods [5][6] and maximum likelihood methods [7][8]. Since these high-resolution methods incorporate the benefits of temporal averaging and knowledge of the model order, the target must be assumed stationary over the period of observation. Thus, these methods fail or suffer performance degradations when the DOAs of the target exhibits significant motion during the observation period.

In recent years there have been several methods developed for estimating or tracking the DOAs of moving targets using passive sensors or arrays of sensors , e.g., [9][10][11], etc. Like the high-resolution methods, these approaches also assume the targets are stationary over a limited time interval. The approach in [10] is based on adaptively estimating a noise subspace basis from the received signal covariance matrix. These methods then rely on a high-resolution technique such as MUSIC [5] to estimated the desired DOAs. In [9], a method based on maximum likelihood estimation of a novel state-space representation for tracking is presented.

An important consideration in target tracking problems is the data association problem; i.e., the association of tracks with measurements. In the case where passive arrays of sensors are used, the data association problem reduces to the association of targets before and after their DOA tracks cross each other. In [11], a method for DOA tracking for disparately-spaced sensors using the EM algorithm is presented. This method treats the DOAs as unknown parameters and the data associations as the missing data.

In this paper, we discuss the use of sequential MC (Monte Carlo) methods for target tracking. MC and MCMC (Markov chain Monte Carlo) methods [12] [13] have been capturing the attention of researchers in the field of statistics throughout the past decade and have more recently emerged

as useful methods in the signal processing arena. They are Bayesian methods based on the idea of numerically sampling posterior distributions of interest that are difficult or impossible to handle analytically. Statistical inferences on parameters of interest can then be made from the resulting histograms. However, conventional MC methods are not well suited to problems where data arrive sequentially, due to excessive computational requirements. This consideration has motivated the development of sequential MC methods (also known as *particle filters*) [14][15], which are capable of recursively updating the probability distributions of interest as new data become available.

In this paper, we propose the application of particle filters to joint detection, estimation and tracking of an unknown and time–varying number of sources. There are several advantages offered by this approach. Firstly, previous methods require prior determination of model order. The MDL and AIC criteria [16] are often used for this purpose. These methods require the assumption of stationarity and are highly sensitive to the white noise assumption. The proposed approach offers robust estimation of the model order *jointly* with other parameters of interest, and furthermore can accommodate changes in model order occurring arbitrarily throughout the observation interval. Secondly, the particle filtering approach estimates the posterior distribution of the parameters given all past data. This distribution can then be marginalized to yield the "instantaneous" posterior distribution of the desired parameters at the current time instant. Thus we need not assume stationarity. This is in contrast to most other methods which involve estimation of second- or higher-order statistics by temporal averaging, a process which requires stationarity over an appropriate interval. Thirdly, with the particle filtering approach, the joint posterior distribution of the target amplitudes given the received data is readily available. This greatly facilitates high-accuracy data association. Finally, in contrast to other methods, because any form of MCMC technique produces an approximation to the entire distribution of interest, one can easily calculate confidence intervals, marginalize with respect to desired parameters, or make inferences on the parameters, etc.

One of the difficulties with particle filtering is the loss of statistical diversity in the recursive update of the importance weights [14]. To mitigate this difficulty, we have introduced a new form of the Reversible Jump MCMC [17][18] process as a novel resampling engine. We propose the use of two new moves, called the *split/merge* moves, which are specifically designed to handle the crossing of the signal tracks, and allow for the joint detection and tracking of the number of

sources.

The paper is organized as follows. Section II presents the state-space model. In sections III and IV, we discuss and extend the particle filtering approach as developed in [14] to the target tracking problem. Results from simulations are presented in Section V where our results are compared with the method presented in [10] [2]. Conclusions are given in Section VI.

Notation: Bold upper case symbols denote matrices, bold lower case symbols denote vectors. The superscript $'$ denotes the Hermitian transpose operation, and the symbol "$\sim$" means "distributed as". We make use of the following probability density functions:

- $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote a complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$p(\boldsymbol{x}) = \frac{1}{|\pi \boldsymbol{\Sigma}|} e^{-(\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{1}$$

- U is a uniform distribution over $[0, 1]$.

$$p(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

- $\mathcal{IG}(\nu, \gamma)$ is an inverse Gamma distribution with parameters ν and γ.

$$p(x) \propto x^{-(\nu+1)} e^{-\gamma/x}. \tag{3}$$

II. The State-Space Model

The problem of interest is the sequential detection of the number of sources impinging an array and the estimation of their corresponding directions of arrival. For example, for a uniform linear array composed of M sensors with known geometry, the manifold is described by the steering matrix, $\boldsymbol{S}(\boldsymbol{\phi}(t)) \in \mathcal{C}^{M \times k(t)}$, the kth column of which is:

$$\boldsymbol{s}(\phi_k(t)) = [1, e^{j\omega_0 d_0 \sin(\phi_k(t))/v}, e^{j2\omega_0 d_0 \sin(\phi_k(t)/v)}, \ldots, e^{j(M-1)\omega_0 d_0 \sin(\phi_k(t))/v}]^T. \tag{4}$$

where $k = 1, \ldots, k(t)$, and $k(t)$ is the model order (number of sources) at time t. This notation is adopted for $k(t)$ to emphasize that model order is unknown and may change at any time throughout the observation interval. In particular, we assume $d_0 \leq \frac{\pi v}{\omega_0}$, where d_0 is the distance between the sensors, v is the velocity of propagation and ω_0 is the frequency of interest.

The signal model we consider consists of a complex vector of observations $\boldsymbol{y}(t) \in \mathcal{C}^M$ which represents the data received by an array of sensors at the tth snapshot. The observation vector

is composed of incident narrow-band plane wave signals from $k(t)$ distinct sources embedded in Gaussian noise. Each of these incident plane-wave signals impinges on the array of sensors at a physical angle $\phi_k, k = 1, \ldots, k(t)$, relative to the normal of the array. The amplitudes of the sources at the tth time instant are denoted by the vector $\boldsymbol{a}(t) \in \mathcal{C}^{k(t)}$.

The sequential sampling approach we adopt admits a first order state-space hidden Markov model. The states $[\boldsymbol{\phi}(t), \boldsymbol{a}(t)]$ evolve according to:

$$\boldsymbol{\phi}(t) = \boldsymbol{\phi}(t - 1) + \sigma_v \boldsymbol{v}(t), \tag{5}$$

$$\boldsymbol{a}(t) \sim \mathcal{N}(\boldsymbol{0}, \sigma_a^2 \boldsymbol{I}_{k(t)}), \tag{6}$$

whereas the observation is given by:

$$\boldsymbol{y}(t) = \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t) + \sigma_w \boldsymbol{w}(t). \tag{7}$$

The noise variables $\boldsymbol{v}(t)$, $\boldsymbol{w}(t) \in \mathcal{C}^M$ are *iid* Gaussian variables with zero mean and unit variance, independent of the parameters. The respective variances of the scaled noise terms are σ_v^2 and σ_w^2. The variance of the amplitudes σ_a^2 satisfies $\sigma_a^2 = \delta^2 \sigma_w^2$, where the hyperparameter δ^2 is set to an a priori estimate of the SNR [18]. The dimension $k(t)$ of the model is described by the following stochastic relationship at time t:

$$k(t) = k(t - 1) + \epsilon_k(t), \tag{8}$$

where the $\epsilon_k(t)$ are discrete *iid* random variables such that

$$P(\epsilon_k(t) = -1) = h/2$$

$$P(\epsilon_k(t) = 0) = 1 - h$$

$$P(\epsilon_k(t) = 1) = h/2, \tag{9}$$

where $h \in [0, 1]$. In (9), it is tacitly assumed that the model order changes by no more than one in each sample period.

In the proposed system of equations, the noise variances σ_v^2 and σ_w^2 are assumed unknown but constant over time. The unknown vectors of amplitudes $\boldsymbol{a}(t)$ are assumed *iid* between snapshots.

We introduce a vector $\boldsymbol{\theta}$ of all the parameters describing the model:

$$\boldsymbol{\theta}_{1:t} \triangleq (\{\boldsymbol{\phi}_{k(t)}\}_{1:t}, \{\boldsymbol{a}_{k(t)}\}_{1:t}, k_{1:t}, \sigma_v^2, \sigma_w^2), \tag{10}$$

where the notation $(\cdot)_{1:t}$ indicates all the elements from time 1 to time t, and the subscript $k(t)$ indicates the size of the corresponding vector. The posterior distribution of interest is then given by $\pi(\boldsymbol{\theta}_{1:t}) \triangleq p(\boldsymbol{\theta}_{1:t}|\boldsymbol{y}_{1:t})$, and can be specified within a normalizing constant, using Bayes' theorem, as:

$$\pi(\boldsymbol{\theta}_{1:t}) \propto p(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}_{1:t})p(\boldsymbol{\theta}_{1:t}) \tag{11}$$

where $p(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}_{1:t})$ is the likelihood function and $p(\boldsymbol{\theta}_{1:t})$ is the prior distribution of the parameters. From the model description, it is clear that the prior distributions for some of the parameters are conditional on $k_{1:t}$ and also on other parameters. Thus, we expand the posterior distribution in (11) to give:

$$\pi(\boldsymbol{\theta}_{1:t}) \propto p(\boldsymbol{y}_{1:t}|\boldsymbol{\phi}_{1:t}, \boldsymbol{a}_{1:t}, k_{1:t}, \sigma_v^2, \sigma_w^2) \times$$
$$p(\boldsymbol{\phi}_{1:t}|\sigma_v^2, k_{1:t})p(\boldsymbol{a}_{1:t}|\boldsymbol{\phi}_{1:t}, \sigma_w^2, k_{1:t})p(k_{1:t})p(\sigma_v^2)p(\sigma_w^2). \tag{12}$$

We now assign distributions for each of the terms in (12). It is assumed that the observations, given the states, are *iid* and that the conditional update likelihoods of the states are also *iid*. Therefore, assuming the distribution of the initial states to be uniform, and using the Markov properties of the model, the distributions of (12) can be written in the form:

$$p(\boldsymbol{y}_{1:t}|\boldsymbol{\phi}_{1:t}, \boldsymbol{a}_{1:t}, k_{1:t}, \sigma_w^2) = \prod_{l=1}^{t} \mathcal{N}(\boldsymbol{S}(\boldsymbol{\phi}_l)\boldsymbol{a}_l, \sigma_w^2 \boldsymbol{I}_M), \tag{13}$$

$$p(\boldsymbol{\phi}_{1:t}|k_{1:t}, \sigma_v^2) = \prod_{l=1}^{t} \mathcal{N}(\boldsymbol{\phi}_{l-1}, \sigma_v^2 \boldsymbol{I}_{k_l}), \tag{14}$$

$$p(\boldsymbol{a}_{1:t}|\boldsymbol{\phi}_{1:t}, k_{1:t}, \sigma_w^2) = \prod_{l=1}^{t} \mathcal{N}(\boldsymbol{0}, \delta^2 \sigma_w^2 (\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l))^{-1}), \tag{15}$$

$$p(k_{1:t}) = \prod_{l=1}^{t} p(k_l|k_{l-1}) = \prod_{l=1}^{t} \epsilon_k(l). \tag{16}$$

The prior distribution on the variances σ_v^2 and σ_w^2 are both assumed to follow the inverse Gamma distribution, which is the conjugate distribution for the Normal distribtuion:

$$p(\sigma_v^2) \sim \mathcal{IG}(\frac{\nu_0}{2}, \frac{\gamma_0}{2}), \tag{17}$$

$$p(\sigma_w^2) \sim \mathcal{IG}(\nu_1, \gamma_1). \tag{18}$$

Since $p(\sigma_v^2)$ is later combined with a real Normal distribution (instead of complex), the factors of $\frac{1}{2}$ in (17) are required to maintain the conjugate property. The above priors are noninformative when the hyperparameters ν and γ are set to zero.

The model is now clearly defined. In the application addressed in this paper, the parameters of interest are primarily the DOAs $\phi_{1:t}$ and the model order $k_{1:t}$. The amplitudes $\boldsymbol{a}_{1:t}$, along with the state update noise variances σ_w^2 and σ_v^2, may be considered nuisance parameters. Even though it is straightforward to numerically marginalize the posterior density to eliminate these undesired parameters using the proposed Monte-Carlo based estimation methods, the resulting procedure is more efficient if the nuisance parameters can be integrated out analytically. Such is the case with the signal amplitudes. We now proceed to eliminate the amplitudes from the posterior distribution $\pi(\boldsymbol{\theta}_{1:t})$ by marginalization.

Using the *iid* Normal distribution of the noise variables, and the model structure given by (5) to (7), and equations (13) to (18), the posterior distribution $\pi(\boldsymbol{\theta}_{1:t})$ of (12) can be written as

$$
\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\theta}_{1:t}) \propto & \prod_{l=1}^{t} \frac{1}{\sigma_w^{2M} \pi^M} \exp\left[-\frac{1}{\sigma_w^2}(\boldsymbol{y}_l - \boldsymbol{S}(\boldsymbol{\phi}_l)\boldsymbol{a}_l)'(\boldsymbol{y}_l - \boldsymbol{S}(\boldsymbol{\phi}_l)\boldsymbol{a}_l)\right] \\
& \times \prod_{l=1}^{t} \frac{|\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l)|}{(\delta\sigma_w)^{2k_l}\pi^{k_l}} \exp\left[-\frac{1}{\delta^2\sigma_w^2}(\boldsymbol{a}_l)'(\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l))(\boldsymbol{a}_l)\right] \\
& \times \prod_{l=1}^{t} \frac{1}{\sigma_v^{2k_l/2}(2\pi)^{k_l/2}} \exp\left[-\frac{1}{2\sigma_v^2}(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})'(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})\right] \qquad (19) \\
& \times \sigma_v^{2(-\frac{\alpha_o}{2}-1)} \exp\left[\frac{-\gamma_o}{2\sigma_v^2}\right] \times \sigma_w^{2(-\alpha_1-1)} \exp\left[\frac{-\gamma_1}{\sigma_w^2}\right] \\
& \times \prod_{l=1}^{t} p(k_l|k_{l-1}),
\end{aligned}
$$

The terms relating to the amplitudes can be collected together to give the following expression,

as in [18]:

$$\boldsymbol{\pi}(\boldsymbol{\theta}_{1:t}) \propto \prod_{l=1}^{t} \frac{1}{\sigma_w^{2^{k_l}} \pi^{k_l}} \exp\left[\frac{-1}{\sigma_w^2}(\boldsymbol{a}_l - \boldsymbol{m}_{\boldsymbol{a}_l})' \boldsymbol{\Sigma}_{k_l}^{-1}(\boldsymbol{a}_l - \boldsymbol{m}_{\boldsymbol{a}_l})\right]$$

$$\times \prod_{l=1}^{t} \frac{1}{\sigma_w^{2^M} \delta^{2^{k_l}}} \exp\left[\frac{-1}{\sigma_w^2} \boldsymbol{y}_l' \boldsymbol{P}_s^{\perp}(\boldsymbol{\phi}_l) \boldsymbol{y}_l\right]$$

$$\times \prod_{l=1}^{t} \frac{|\boldsymbol{S}'(\boldsymbol{\phi}_l) \boldsymbol{S}(\boldsymbol{\phi}_l)|}{\sigma_v^{2^{(k_l/2)}}(2\pi)^{(k_l/2)}} \exp\left[\frac{-1}{2\sigma_v^2}(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})'(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})\right] \quad (20)$$

$$\times \sigma_v^{2^{(-\frac{\nu_o}{2}-1)}} \exp\left[\frac{-\gamma_o}{2\sigma_v^2}\right] \times \sigma_w^{2^{(-\nu_1-1)}} \exp\left[\frac{-\gamma_1}{\sigma_w^2}\right]$$

$$\times \prod_{l=1}^{t} p(k_l|k_{l-1}),$$

where

$$\boldsymbol{\Sigma}_{k_l}^{-1} = \boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l)(1 + \delta^{-2}),$$

$$\boldsymbol{m}_{\boldsymbol{a}_l} = \boldsymbol{\Sigma}_{k_l} \boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{y}_l \quad (21)$$

and

$$\boldsymbol{P}_s^{\perp}(\boldsymbol{\phi}_l) = \boldsymbol{I}_M - \frac{\boldsymbol{S}(\boldsymbol{\phi}_l)(\boldsymbol{S}'(\boldsymbol{\phi}_l)\boldsymbol{S}(\boldsymbol{\phi}_l))^{-1}\boldsymbol{S}'(\boldsymbol{\phi}_l)}{(1 + 1/\delta^2)}. \quad (22)$$

From (20) a *maximum a posteriori* estimate of the amplitudes, knowing the other parameters is readily available as:

$$\hat{\boldsymbol{a}}_{MAP}(l) = \boldsymbol{m}_{\boldsymbol{a}_l}. \quad (23)$$

Thus, the amplitude parameters need not be included in the particle filter. Instead, they can be estimated at each iteration, after the sampling of the other parameters as discussed in Sects. 3 and 4.

It then becomes straightforward to integrate out the amplitudes in (20) to yield a simpler definition of the posterior distribution in terms of the remaining parameters. The posterior

distribution can then be simplified to:

$$
\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\phi}_{1:t}, \sigma_v^2, \sigma_w^2, k_{1:t}) \propto & \prod_{l=1}^{t} \frac{1}{\sigma_w^{2M}(1+\delta^2)^{k_l}} \exp\left[\frac{-1}{\sigma_w^2} \boldsymbol{y}_l' \boldsymbol{P}_s^{\perp}(\boldsymbol{\phi}_l) \boldsymbol{y}_l\right] \\
& \times \prod_{l=1}^{t} \frac{1}{\sigma_v^{2^{(k_l/2)}}(2\pi)^{(k_l/2)}} \exp\left[\frac{-1}{2\sigma_v^2}(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})'(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})\right] \\
& \times \sigma_v^{2^{(-\frac{\nu_o}{2}-1)}} \exp\left[\frac{-\gamma_o}{2\sigma_v^2}\right] \times \sigma_w^{2^{(-\nu_1-1)}} \exp\left[\frac{-\gamma_1}{\sigma_w^2}\right] \\
& \times \prod_{l=1}^{t} p(k_l|k_{l-1}).
\end{aligned}
\tag{24}
$$

The MAP estimators of the nuisance parameters of the variances can be readily obtained by comparing the previous distribution with a product of inverse Gamma distributions. Using the fact that the mode of the inverse Gamma distribution is $\frac{\gamma}{\nu+1}$, it follows that

$$
\sigma_{v_{MAP}}^2(t) = \frac{\frac{\gamma_0}{2} + \frac{1}{2}\sum_{l=1}^{t}(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})'(\boldsymbol{\phi}_l - \boldsymbol{\phi}_{l-1})}{\frac{\nu_0}{2} + \frac{1}{2}\sum_{l=1}^{t} k(l) + 1},
\tag{25}
$$

$$
\sigma_{w_{MAP}}^2(t) = \frac{\gamma_1 + \sum_{l=1}^{t}(\boldsymbol{y}_l \boldsymbol{P}_s^{\perp}(\boldsymbol{\phi}_l)\boldsymbol{y}_l)'}{\nu_1 + Mt + 1}.
\tag{26}
$$

We choose however to keep these parameters in the expression of the posterior distribution in (24). This simplifies the derivation of the acceptance probabilities of the moves as discussed in Sect. 4. Since the nuisance parameters can be estimated, we now define a new vector $\boldsymbol{\alpha}$ of parameters to sample, as:

$$
\boldsymbol{\alpha}_{1:t} \stackrel{\Delta}{=} \left(\boldsymbol{\phi}_{1:t}, k_{1:t}\right).
\tag{27}
$$

III. Sequential Importance Sampling

This section briefly describes the SIS procedure, which is used to extract the DOA estimates for tracking. In this paper, the background treatment on the SIS methodology is necessarily brief. The reader is referred to [14] [19] [25] and the references therein for a more complete coverage of this topic. In this section, we first describe the Bayesian importance sampling procedure. We then outline the recursive procedure to update the desired histogram from time $t-1$ to time t, without the need for recomputing the entire joint distribution all over again as new data become available.

We now describe the Bayesian importance sampling scheme [20] in a general framework. Our objective is to generate a numerical approximation of an arbitrary distribution $p(\boldsymbol{x})$, in the form

of a histogram, by drawing a large number of samples from $p(\boldsymbol{x})$. However, in many practical cases, it is not easy to sample directly from this distribution, since it may be in a "non-standard" form, multivariate, and known only up to a normalizing constant. Therefore, in the Bayesian importance sampling paradigm, N samples $\boldsymbol{x}^{(i)}, i = 1, \ldots, N$ (particles) are drawn from another "easy-to-sample-from" function $q(\boldsymbol{x})$ called the "importance function", whose support includes that of $p(\boldsymbol{x})$. The histogram of these samples approximates the distribution $q(\boldsymbol{x})$. In order to transform these samples to represent the desired distribution, we form the histogram $\hat{p}_N(d\boldsymbol{x})$ approximating $p(\boldsymbol{x})$ as:

$$\hat{p}_N(d\boldsymbol{x}) = \frac{\sum_{i=1}^N \tilde{w}(\boldsymbol{x}^{(i)})\delta_{\boldsymbol{x}^{(i)}}(d\boldsymbol{x})}{\sum_{i=1}^N \tilde{w}(\boldsymbol{x}^{(i)})}. \tag{28}$$

where

$$\tilde{w}(\boldsymbol{x}^{(i)}) = \frac{p(\boldsymbol{x}^{(i)})}{q(\boldsymbol{x}^{(i)})}, \tag{29}$$

and $d\boldsymbol{x}$ is a small, finite region surrounding an $\boldsymbol{x}$ of interest (i.e., a histogram "bin"), and $\delta_{\boldsymbol{x}^{(i)}}(d\boldsymbol{x})$ is a Dirac delta defined as:

$$\delta_{\boldsymbol{x}^{(i)}}(d\boldsymbol{x}) = \begin{cases} 1, \text{if } \boldsymbol{x}^{(i)} \in d\boldsymbol{x} \\ 0, \text{ otherwise.} \end{cases} \tag{30}$$

Since $p(\boldsymbol{x})$ in (29) can only be easily determined up to a normalizing constant, the denominator of (28) in effect normalizes the weights $\tilde{w}(\boldsymbol{x}^{(i)})$ so that $\hat{p}_N(d\boldsymbol{x})$ is a proper distribution. It can be shown that the estimate $\hat{p}_N(\boldsymbol{x})$ is biased; however, under mild conditions, the expectation $\bar{I}_N(f(t, \cdot)$ of any function $f(t, \boldsymbol{x})$ over $\hat{p}_N(\boldsymbol{x})$ converges to $I(f(t, \cdot))$, which is the expectation of $f(t, \boldsymbol{x})$ over $p(\boldsymbol{x})$, in the sense that:

$$\sqrt{N}\left[\bar{I}_N(f(t, \cdot) - I(f(t, \cdot))\right] \xrightarrow[N \to \infty]{} \mathcal{N}(\boldsymbol{0}, \bar{\sigma}^2_{f(t, \cdot)}), \tag{31}$$

where $\Longrightarrow$ denotes convergence in distribution and $\bar{\sigma}^2_{f(t, \cdot)}$ is the variance of $f(t, \cdot)$ over $\pi(\boldsymbol{\alpha})$.

We now proceed to show how the Bayesian importance sampling approximation can be recursively updated as new data arrives, in order to keep the previously simulated trajectories of the particles, and to avoid sampling the increasingly long parameter vector as a whole. Returning to the DOA tracking problem, we now assume at the $(t - 1)$th time instant a set of weights $\tilde{w}^{(i)}_{1:t-1}, i = 1, \ldots, N$ is available from which the approximation $\hat{\pi}_N(d\boldsymbol{\alpha}_{1:t-1})$ of the joint

distribution $\pi(\boldsymbol{\alpha}_{1:t-1})$ can be determined, so that:

$$\hat{\pi}_N(d\boldsymbol{\alpha}_{1:t-1}) = \frac{\sum_{i=1}^{N} \tilde{w}^{(i)}(\boldsymbol{\alpha}_{1:t-1}^{(i)})\delta_{\boldsymbol{\alpha}_{1:t-1}^{(i)}}(d\boldsymbol{\alpha}_{1:t-1})}{\sum_{i=1}^{N} \tilde{w}^{(i)}(\boldsymbol{\alpha}_{1:t-1}^{(i)})}. \tag{32}$$

The joint posterior distribution of all parameters from time 1 to t can be written using Bayes' theorem as:

$$\pi(\boldsymbol{\alpha}_{1:t}) = \frac{p(\boldsymbol{y}_{1:t}|\boldsymbol{\alpha}_{1:t})p(\boldsymbol{\alpha}_{1:t})}{p(\boldsymbol{y}_{1:t})}. \tag{33}$$

It can be shown, using the Markov properties of the model and the *iid* assumptions on the noise variables, that (33) can be written in the recursive, time-update form:

$$\pi(\boldsymbol{\alpha}_{1:t}) = \pi(\boldsymbol{\alpha}_{1:t-1}) \times \frac{p(\boldsymbol{y}_t|\boldsymbol{\alpha}_t)p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1})}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})}. \tag{34}$$

In principle, this recursion would allow us to sequentially and recursively compute the posterior distribution. However, it is not useful in its present form, since the normalizing constant $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})$ and the desired marginal distributions require the evaluation of complex, multi-dimensional integrals, which are generally difficult or impossible to evaluate analytically. We therefore resort to the Bayesian importance sampling scheme.

Since the importance function may be chosen at our discretion, it can be selected to obey the following property:

$$q(\boldsymbol{\alpha}_{1:t}|\boldsymbol{y}_{1:t}) = q(\boldsymbol{\alpha}_{1:t-1}|\boldsymbol{y}_{1:t-1})q(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{1:t-1},\boldsymbol{y}_{1:t}). \tag{35}$$

Equations (34) and (35) then define a recursion on the weights of the ith particle at time t:

$$\tilde{w}^{(i)}(t) = w^{(i)}(t-1)\frac{p(\boldsymbol{y}_t|\boldsymbol{\alpha}_t^{(i)})p(\boldsymbol{\alpha}_t^{(i)}|\boldsymbol{\alpha}_{t-1}^{(i)})}{q(\boldsymbol{\alpha}_t^{(i)}|\boldsymbol{\alpha}_{1:t-1}^{(i)},\boldsymbol{y}_{1:t})} \quad i = 1,\ldots,N. \tag{36}$$

The recursion is made complete with $\tilde{w}^{(i)}(1)$ being defined as:

$$\tilde{w}^{(i)}(1) = \frac{\pi(\boldsymbol{\alpha}^{(i)}(1)|\boldsymbol{y}(1))}{q(\boldsymbol{\alpha}^{(i)}(1)|\boldsymbol{y}(1))}. \tag{37}$$

Eq. (35) and (37) represent the desired recursion of the weights. The method proceeds by sampling N particles $\boldsymbol{\alpha}^{(i)}(t), i = 1,\ldots,N$ from the importance function $q(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{1:t-1},\boldsymbol{y}_{1:t})$ in (35). The terms $p(\boldsymbol{y}_t|\boldsymbol{\alpha}_t)$ and $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1})$ can then be evaluated from these particles using the model (13) and (14), and the new data $\boldsymbol{y}(t)$. Note that in (36) we have omitted the unknown

normalizing component $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})$. It is straightforward to show [14] that the effect of this omission is compensated by normalizating the weights, as follows:

$$w^{(i)}(t) = \frac{\tilde{w}^{(i)}(t)}{\sum_{i=1}^{N} \tilde{w}^{(i)}(t)}. \tag{38}$$

A major difficulty with SIS methods in general is that in practice, the recursion of (36) degenerates quickly, even after a few iterations, so that only a handful of significantly-valued particles remain. Therefore, any estimate based on these very few particles would show a large variance. In Sect. 4 we discuss methods to mitigate this effect.

At each observation time t, samples from the importance function must be generated. It is shown in [21], that the optimal importance function that satisfies the recurrence requirement (35) and minimizes the variance of the weights generated by the recursion (36), is given by:

$$q_{optimal}(\cdot) = q(\boldsymbol{\alpha}^{(i)}(t)|\boldsymbol{\alpha}^{(i)}(t-1), \boldsymbol{y}(t)). \tag{39}$$

This distribution is not easily evaluated directly for the problem at hand. However, an approximation is readily obtained by means of a local linearization (Taylor expansion) of the observation equation. The observation equation (7) and the state update equation (5) for $\boldsymbol{\phi}$ are reproduced here for convenience:

$$\boldsymbol{\phi}(t) = \boldsymbol{\phi}(t-1) + \sigma_v \boldsymbol{v}(t),$$

$$\boldsymbol{y}(t) = \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t) + \sigma_w \boldsymbol{w}(t),$$

These equations yield:

$$\boldsymbol{y}(t) \approx \boldsymbol{S}(\boldsymbol{\phi}(t-1))\boldsymbol{a}(t) + \left.\frac{\partial \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t)}{\partial \boldsymbol{\phi}(t)}\right|_{\left(\begin{array}{c}\boldsymbol{\phi}(t) = \boldsymbol{\phi}(t-1)\\ \boldsymbol{a}(t) = \boldsymbol{a}(t-1)\end{array}\right)} \times \left(\boldsymbol{\phi}(t) - \boldsymbol{\phi}(t-1)\right) + \sigma_w \boldsymbol{w}(t). \tag{40}$$

After solving the above for $\boldsymbol{\phi}(t)$, from the assumptions on the model noise, we see the resulting distribution for $\boldsymbol{\phi}(t)\big|\boldsymbol{\phi}(t-1), \boldsymbol{y}(t)$, i.e., the optimal importance function, is linear and Gaussian and can be expressed as:

$$q(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), \boldsymbol{y}(t)) \sim \mathcal{N}(\boldsymbol{m}(t), \boldsymbol{\Sigma}(t)), \tag{41}$$

where, for each particle:

$$\boldsymbol{\Sigma}^{-1}(t) = \sigma_v^{-2}(t)\boldsymbol{I}_k(t) + \boldsymbol{G}^{-1}(\sigma_w^{-2}(t)\boldsymbol{I}_M))\boldsymbol{G}, \tag{42}$$

$$\boldsymbol{m}(t) = \boldsymbol{\Sigma}(t)\left(\sigma_v^{-2}(t)\boldsymbol{I}_k(t) + \boldsymbol{G}^{-1}(\sigma_w^{-2}(t)\boldsymbol{I}_M)[\boldsymbol{y}(t) - \boldsymbol{S}(\boldsymbol{\phi}(t-1))\boldsymbol{a}(t-1) + \boldsymbol{G}\boldsymbol{\phi}(t-1)]\right), \quad (43)$$

where the matrix $\boldsymbol{G}$ is the gradient of the observation equation:

$$\boldsymbol{G} = \frac{\partial \boldsymbol{S}(\boldsymbol{\phi}(t))\boldsymbol{a}(t)}{\partial \boldsymbol{\phi}(t)}. \tag{44}$$

In summary, the recursive update for the weights is obtained using the following form of (36):

$$\tilde{w}^{(i)}(t) = w^{(i)}(t-1) \times \frac{p\big(\boldsymbol{y}(t)|\boldsymbol{\phi}^{(i)}(t), k^{(i)}(t), \boldsymbol{a}^{(i)}(t), \sigma_w^{2^{(i)}}\big)p\big(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), k^{(i)}(t), \sigma_v^{2^{(i)}}\big)}{q(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), \boldsymbol{y}(t))}.$$

$$(45)$$

Eqs. (13) and (14) are used for the respective terms on the numerator, and (41) is used for the denominator.

Even though the importance function (41) is chosen carefully, in practice the algorithm still quickly degenerates. It is thus necessary to resample the particles according to their importance weights. This operation can be done very efficiently, with $\mathcal{O}(N)$ operations [22]. However, the trajectories with high importance weights are statistically selected many times, limiting the true statistical diversity amongst the particles.

A more clever approach, [1][21] [26] uses a reversible jump MCMC [17][18] step on each particle at time t. The MCMC procedure samples $\pi(\boldsymbol{\alpha})$ directly, thus introducing statistical diversity amongst the particles. Also, the reversible jump process is capable of exploring parameter spaces of varying dimension, which as we see in Sect. 4 is the key to detection of model order [18][23].

The above procedure is summarized in the following schema:

Sequential Importance Sampling for Tracking an Unknown Number of DOA's

For time $t = 1$, initialize the weights $w^{(i)}(1), i = 1, \ldots, N$ using (37).

For each time step $t = 2, 3, \ldots$, DO

1. The Importance Sampling Step:

   • For $i = 1, \ldots, N$, generate the particles by sampling from the distribution $q(\cdot|\cdot)$, as follows (see (41)):

$$q(\boldsymbol{\phi}^{(i)}(t)|\boldsymbol{\phi}^{(i)}(t-1), \boldsymbol{y}(t)) \sim \mathcal{N}(\boldsymbol{m}(t), \boldsymbol{\Sigma}(t)) \tag{46}$$

   • For $i = 1, \ldots, N$, Evaluate the un-normalized importance weights from (45).

- For $i = 1, \ldots, N$, normalize the weights:

$$w^{(i)}(t) = \frac{\tilde{w}^{(i)}(t)}{\sum_{j=1}^{N} \tilde{w}^{(j)}(t)} \tag{47}$$

2. The Resampling/Selection of the Particles:

   - Sample a vector of index l distributed as:

$$P(l(j) = i) = w^{(i)}(t) \tag{48}$$

   - Resample the particles with the index vector:

$$\phi_{0:k}^{(i)} = \phi_{0:k}^{(l(i))} \tag{49}$$

   - Re-assign all the weights to $w^{(i)}(t) = \frac{1}{N}$.

3. The Reversible Jump MCMC Step

   - Apply the sampler to be described in section IV to enhance diversity amongst the particles and facilitate detection of model order.

   ∎

The SIS procedure is now completely described. In order to use this procedure to track the DOAs, our objective is to estimate the parameters of interest $\phi(t)$ and $k(t)$ given all past observations, at each time instant. This can be achieved by forming the marginal distribution corresponding only to the specific parameters of interest from $\pi(\boldsymbol{\alpha}_{1:t})$. One of the primary advantages of using a numerical Bayesian procedure for parameter estimation is that this implicit integration is readily performed directly from the histogram $\hat{\pi}_N(d\boldsymbol{\alpha}_{1:t})$.

IV. The Reversible Jump MCMC Diversity Step

The reversible jump MCMC process is a variation of the Metropolis-Hastings (MH) algorithm [20]. The algorithm inherently sets up a Markov chain whose invariant distribution corresponds to the posterior of interest. After an appropriate "burn-in" period which is required for the Markov chain to reach equilibrium, the states at successive iterations represent samples from the distribution of interest.

In this application, we use the MH method to sample the posterior distribution with respect to $\phi(t)$ and $k(t)$. Assume at the jth iteration of the chain we are in state $(\phi^{(j)}, k^{(j)})$. A candidate $(\phi^{\star}, k^{\star})$ for the next state of the chain is drawn at random from a proposal distribution $d(\cdot|\cdot)$,

which may be conditional on $(\phi^{(j)}, k^{(j)})$. An acceptance ratio r is then generated according to:

$$r = \frac{\pi(\phi^{\star}, k^{\star})d(\phi^{(j)}, k^{(j)}|\phi^{\star}, k^{\star})}{\pi(\phi^{(j)}, k^{(j)})d(\phi^{\star}, k^{\star}|\phi^{(j)}, k^{(j)})}J, \tag{50}$$

where J is the Jacobian of the transformation from ϕ to $\phi^{\star}$. An acceptance parameter ξ is then defined as:

$$\xi = \min\{r, 1\}. \tag{51}$$

Then, the proposed candidate $(\phi^{\star}, k^{\star})$ is accepted as the current state at iteration $j + 1$ with probability ξ.

In our application, since the dimension of the parameter space $\phi_{1:t}$ varies with $k_{1:t}$, we use the *reversible jump* MCMC method which samples directly from the joint distribution over all model orders of interest. In effect, the process jumps between subspaces of different dimensions, thus visiting all relevant model orders. In the reversible jump case, candidate samples are chosen from a *set* of proposal distributions, which are randomly accepted according to an acceptance ratio that ensures reversibility, and therefore the invariance of the Markov chain with respect to the desired posterior distribution. Here, we choose our set of proposal distributions to correspond to the following set of moves:

• the *birth move*, chosen with probability b_k, for which a new source is proposed at random; i.e., $k(t) = k(t - 1) + 1$.

• the *death move*, chosen with probability d_k, for which one of the existing sources is proposed to be removed; i.e., $k(t) = k(t - 1) - 1$.

These moves, in conjunction with the update move below, enable us to sample the parameter $k_{1:t}$. By forming the marginal of $\pi(\alpha_{1:t})$ with respect to $k_{1:t}$, we can detect the most likely number of sources vs. time. In addition to these moves, we propose two further novel moves, which we have shown in simulations to improve performance when two neighbouring DOA tracks cross. These additional moves are

• *the split move*, for which an existing source is proposed to be split into two sources. This move is chosen with probability s_k .

• *the merge move*, for which two neighbouring sources are proposed to be merged into one. This move is chosen with probability m_k .

We also have the update move:

- with the *update move*, all the parameters are updated with fixed dimension; i.e., $k(t) = k(t-1)$. This move is executed with probability $1 - b_k - d_k - s_k - m_k$. With the update, birth and death moves, $\boldsymbol{J}$ in (50) is readily shown [18] to be unity.

It is shown in [26] that the proposed MCMC sampling procedure requires no burn in period in this application. This is a consequence of the fact the particles before the MCMC step are already distributed according to the limiting distribution of the chain. Thus, in the interest of computational efficiency, only one MCMC iteration need be applied to each particle at each time step.

Notation: For the following subsections, it is understood that the moves are applied to particle i at time t. Hence, we simplify the notation by omitting reference to both t and i, when no confusion is possible. In order to emphasize that the various moves result in a change in model dimension, we explicitly denote dependence of the respective parameters on model order; e.g, $\boldsymbol{\phi}_k$.

The selection of moves is described by the following schema.

Reversible Jump MCMC

1. Current state of the chain = current state of the particles $(k(t), \boldsymbol{\phi}^{(i)}(t))$.
2. Iteration t for the ith particle, $i = 1, \ldots, N$:
   - Sample $u \sim U$
   - if $(u < b_k)$ then "birth move"
   - else if $(u < b_k + d_k)$ then "death move"
   - else if $(u < b_k + d_k + s_k)$ then "split move"
   - else if $(u < b_k + d_k + s_k + m_k)$ then "merge move"
   - else update all the parameters
3. $t \leftarrow t + 1$, goto step 2

A. Update move

If the update move is selected, all the parameters are resampled, with fixed model order k. The proposal distribution $d\big(\phi_k(t)\big|\phi_k(t-1)\big)$ for the candidate $\phi_k^\star$ is given using (5) as

$$d\big(\phi_k(t)\big|\phi_k(t-1)\big) = \mathcal{N}(\phi_k(t-1), \sigma_v^2(t)\boldsymbol{I}_{k(t)}). \tag{52}$$

By substituting (52) and (24) into (50), after some algebraic manipulation we obtain the following expression for the acceptance ratio for the update move:

$$r_{update} = \frac{\exp^{\frac{-1}{\sigma_w^2}\boldsymbol{y}'(t)\boldsymbol{P}_s^\perp(\phi_{k(t)}^\star)\boldsymbol{y}(t)}}{\exp^{\frac{-1}{\sigma_w^2}\boldsymbol{y}'(t)\boldsymbol{P}_s^\perp(\phi_{k(t)})\boldsymbol{y}(t)}}. \tag{53}$$

The candidate $\phi_k^\star$ is accepted as the ith particle at time t, with probability:

$$\xi_{update} = \min(r_{update}, 1). \tag{54}$$

The amplitude parameters, which are required for data association in section V, are estimated directly from (23). The noise variance parameters σ_w^2 and σ_v^2, which are required in (52) and (53), are estimated directly from (25) and (26).

The update move is summarized with the following schema:

Update Move

- Propose a candidate $\phi^\star$ from (52).
- Evaluate ξ_{update} with (54)
- Sample $u \sim U$
- if $(u \leq \xi_{update})$ then
  - The state of the Markov Chain becomes $(k, \phi^\star)$,
  - If desired, estimate $\boldsymbol{a}|\phi$ from (23)
  - update σ_v^2 and σ_w^2 from (25) and (26) respectively.

else it remains at (k, ϕ)

$\blacksquare$

B. Birth/Death moves

The birth move proposes a candidate in a higher dimension model, as opposed to the death move, which in turn proposes a candidate in a lower dimension model.

For the birth move, a new source ϕ_c is proposed at random from the prior distribution for the directions of arrival:

$$\phi^\star_{k(t)+1} = [\phi_{k(t)} | \phi_c] \tag{55}$$

After straightforward algebra, the acceptance ratio for the birth move is (using (24)):

$$r_{birth} = \frac{\exp^{\frac{-1}{\sigma_w^2} \boldsymbol{y}'(t) \boldsymbol{P}^\perp_s (\phi^\star_{k(t)+1}) \boldsymbol{y}(t)}}{\exp^{\frac{-1}{\sigma_w^2} \boldsymbol{y}'(t) \boldsymbol{P}^\perp_s (\phi_{k(t)}) \boldsymbol{y}(t)}} \times \frac{1}{(1 + \delta^2)(k+1)} \tag{56}$$

with corresponding acceptance probability given by:

$$\xi_{birth} = \min\{r_{birth}, 1\}. \tag{57}$$

If the move is accepted, then the amplitudes and noise variance parameters are then updated in the same manner as described for the update move process.

The death move is just the reverse. A source, amongst the $(k+1)$ sources is randomly selected to be removed. It is straightforward to show the new candidate, of dimension k, is then accepted with probability:

$$\xi_{death} = \min\{\frac{1}{r_{birth}}, 1\}. \tag{58}$$

The schemas for the birth and death moves are similar to that for the update move with appropriate changes. However, for the birth move, if the candidate is accepted, the new state becomes $(k+1, \phi_{k+1})$, otherwise, it remains at (k, ϕ_k). For the death move, if the candidate is accepted, the new state becomes (k, ϕ_k), otherwise, it remains at $(k+1, \phi_{k+1})$.

C. Split/Merge moves

The split move proposes a candidate in a higher dimension model, as opposed to the merge move, which in turn proposes a candidate in a lower dimension model. The split move is designed to handle the situation where two DOA tracks separate after crossing. The merge move corresponds to the case where two adjacent DOA tracks coalesce before crossing.

For the split move, two new sources $\phi^\star_j$ and $\phi^\star_{j+1}$ are proposed as a replacement of the source $\phi_j \in \phi_k$, selected at random amongst the existing k sources:

$$\phi^\star_j = \phi_j - uW, \tag{59}$$

$$\phi^\star_{j+1} = \phi_j + uW, \tag{60}$$

where W is some fixed and known parameter and $u \sim U$. These new angles are inserted in the parameter vector, replacing the jth element, to produce a candidate vector $\phi^{\star}_{k+1}$ for the split move as follows:

$$\phi^{\star}_{k+1} = \left[\phi_k \big(1 : (j-1)\big), \phi^{\star}_j, \phi^{\star}_{j+1}, \phi_k \big((j+1) : k\big) \right]. \tag{61}$$

As opposed to the previously defined moves, the split/merge moves require the evaluation of the Jacobian term in (50). The Jacobian can be evaluated as:

$$J = \left\| \begin{matrix} \frac{\partial u}{\partial \phi_1} & \frac{\partial \phi}{\partial \phi_1} \\ \frac{\partial u}{\partial \phi_2} & \frac{\partial \phi}{\partial \phi_2} \end{matrix} \right\| = \frac{1}{2W}. \tag{62}$$

After some straightforward algebra, the acceptance ratio for the split move is given as:

$$r_{split} = \frac{\exp \left[\frac{-1}{\sigma^2_w} \boldsymbol{y}'(t) \boldsymbol{P}^{\perp}_s (\phi^{\star}_{k(t)+1}) \boldsymbol{y}(t) \right]}{\exp \left[\frac{-1}{\sigma^2_w} \boldsymbol{y}'(t) \boldsymbol{P}^{\perp}_s (\phi_{k(t)}) \boldsymbol{y}(t) \right]}$$

$$\times \frac{\exp \left[\frac{-1}{2\sigma^2_v} \left(\phi^{\star}_{k(t)+1}(t-1) - \phi_{k(t)+1}(t-1) \right)' \left(\phi^{\star}_{k(t)+1}(t-1) - \phi_{k(t)+1}(t-1) \right) \right]}{\exp \left[\frac{-1}{2\sigma^2_v} (\phi_{k(t)}(t) - \phi_{k(t)}(t-1))'(\phi_{k(t)}(t) - \phi_{k(t)}(t-1)) \right]} \tag{63}$$

$$\times \frac{1}{2\sqrt{2\pi}\sigma_v(1+\delta^2)},$$

$$\xi_{split} = \min\{r_{split}, 1\}. \tag{64}$$

This split move is attempted only if no original sources fall between the two proposed candidates, such that the reverse move, the merge move, makes the sampling reversible. This measure is necessary to satisfy the reversibility condition [17], which in turn is sufficient for the invariant distribution of the Markov chain to converge to the desired density.

The merge move is just the reverse of the split move. A source, ϕ_j, amongst the first k of the $(k+1)$ sources is randomly selected. A candidate source $\phi^{\star}_j$ is proposed as the superposition of two adjoining sources:

$$\phi^{\star}_j = (\phi_j + \phi_{j+1})/2. \tag{65}$$

This combined angle is inserted in the parameter vector, replacing elements j and $j+1$ to produce the candidate vector for the merge move as follows:

$$\phi^{\star}_k = \left[\phi_k \big(1 : (j-1)\big), \phi^{\star}_j, \phi_k \big((j+2) : (k+1)\big) \right]. \tag{66}$$

Similar to the death move case, it is straightforward to show that the candidate vector $\phi_k^\star$, of dimension k is accepted with probability:

$$\xi_{merge} = \min\{\frac{1}{r_{split}}, 1\}. \tag{67}$$

For both the split and merge moves, the amplitude and noise variance parameters are also updated in the manner described for the update move.

Split Move

- Pick at random one DOA element ϕ_j, amongst the first k existing directions. Evaluate:

$$\phi_j^\star = \phi_j - uW \tag{68}$$

$$\phi_{j+1}^\star = \phi_j + uW \tag{69}$$

- Evaluate ξ_{split} with (64)
- Sample $u_2 \sim U$
- if $(u_2 \leq \xi_{split})$then
  - The state of the Markov Chain becomes $(k + 1, \phi_{k+1}^\star)$,
  - If desired, estimate $a_{k+1}|k + 1, \phi_{k+1}^\star$ from (23)
  - update σ_v^2 and σ_w^2 from (25) and (26) respectively.

else it remains at (k, ϕ_k)

■

Merge Move

- Pick at random two adjacent directions of arrival among the $(k + 1)$ existing DOA

$$\phi_j^\star = (\phi_j + \phi_{j+1})/2 \tag{70}$$

- Evaluate ξ_{merge} with (58)
- Sample $u \sim U$
- if $(u \leq \xi_{merge})$ then
  - The state of the Markov Chain becomes $(k, \phi_k^\star)$,
  - If desired, estimate $a_k|k, \phi_k^\star$ from (23)
  - update σ_v^2 and σ_w^2 from (25) and (26) respectively.

else it remains at $(k+1, \phi_{k+1})$

$\blacksquare$

V. SIMULATION RESULTS

The proposed algorithm is now verified with simulated data, generated for the true number of sources $k_o = 2$, with parameters described in Table 5.1. The received array is circular[1] and composed of $M = 8$ elements. The parameters and observations evolve according to the state space model (5) - (7), with an initial SNR of 20dB, which is defined as:

$$SNR = \frac{\boldsymbol{a}'(1)\boldsymbol{a}(1)}{\sigma_w^2} \tag{71}$$

Parameter	σ_v^2	σ_w^2	$\phi(0)$	$\boldsymbol{a}(1)$	σ_a^2
Value	5deg^2.	0.15	$[70°, 110°]$	$[2-2j, 4+j]$	0.0707

TABLE 5.1

PARAMETERS OF THE STATE-SPACE MODEL FOR SIMULATED DATA

The hyperparameter $\delta^2 = 100$ in accordance with the initial SNR value, and the hyperparameters $\nu_0, \nu_1, \gamma_0, \gamma_1$ are all set to zero, corresponding to a non-informative prior on the respective variances. (Brief experiments have verified performance is robust to the values of the hyperparameters). The DOAs are simulated using a first–order random walk, with variance σ_v^2, thus generating a mildly nonstationary DOA environment. The parameter k for the model order is initialized at $k(1) = 1$ or $k(1) = k_{max}$ (where k_{max} is the maximum allowable model order) and all the other parameters are initialized at random over their respective parameter space. The initialization is therefore done blindly. The particle filter uses $N = 300$ particles. We compare the performance for the *SIS* method with that of the PASTd method [2] with joint rank estimation.

A. First scenario: Change Point in the number of sources

In this first scenario, the number of sources is initialized to $k = 1$, and is complicated by a change point at $t = 50$, when one of the sources vanishes. Figure 1 shows the results obtained with the particle filter. It shows that the directions of arrival are well traced by their estimates

[1]Circular arrays are used more often in practice as they do not suffer from the ambiguity between the forward and backward look directions, as is inherent to linear arrays.
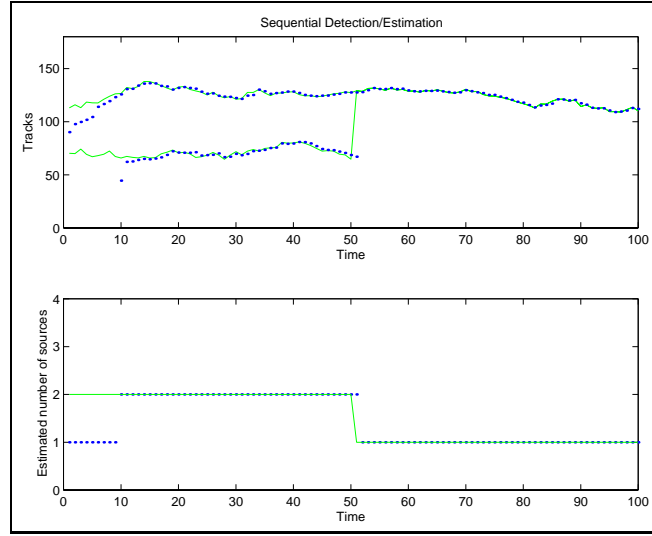
Fig. 1. Top: Sequential estimates of the directions of arrival, and Bottom: the number of detected signals, each vs. time, using the particle filter (Scenario A), initialized to $k(1) = 1$. The finely dotted line shows the true values, and the coarsely dotted line gives the estimated values.

throughout the entire tracking process and that the number of sources is correctly estimated. Initially, since $k = 1$, the algorithm tracks towards only one source, until a birth move is accepted. The second source is then detected and later on correctly estimated. The change-point was detected within one sample period and the estimates of the parameters, following the change point, quickly adapt to the true values. The same scenario is used in figure 2, but the model order is initially set to $k_{max} = 5$. Within 35 observations, the correct model order is determined, and the DOA trajectories follow the true values. Again, the change point is detected within one sample period. It is therefore seen that the proposed particle filter approach performs well under nonstationary conditions and variations in the initial values of the parameters.

We now discuss the comparison of the particle filter results results with those from the PASTd algorithm. Due to the nonstationarity, the PASTd algorithm fails to give meaningful results in this environment. In the nonstationary case, rank estimation fails as the number of sources is always over-estimated and the DOA estimates obtained are smoothed versions of the true values. This behaviour is typical of any algorithm which is based on time-averaged statistics.

We therefore consider a simulation scenario which is more favourable to the PASTd algorithm. In this case, the source DOAs are held steady at $\pm 30°$ for the first 500 observations. As shown on figure 3, the PAST algorithm with joint rank estimation is much slower to converge to the
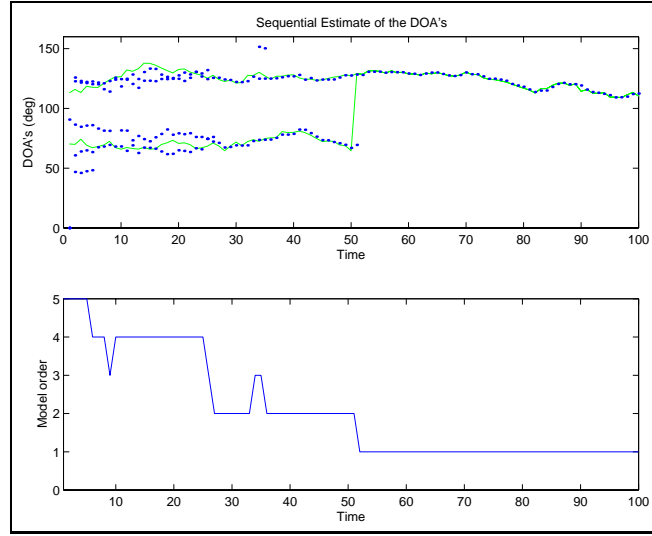
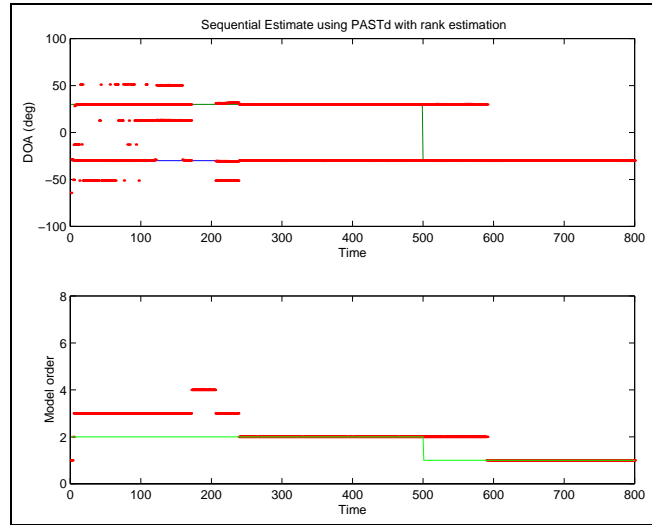Fig. 2. Same as Figure 1, except the filter is initialized to $k(1) = k_{max} = 5$.

Fig. 3. Sequential estimates of directions of arrival using PAST and root-MUSIC (Scenario A).

true number of sources than the particle filter case, both initially and after the change point. For this favourable case, the DOA estimates produced by the algorithm include values which are close to the true values, as shown.

B. Second scenario: Sources crossing

In this subsection, we apply the previously developed algorithm to a scenario where the source DOAs cross, thus verifying the performance of the split/merge move combinations. In this

scenario, the variance σ_v^2 of the update equation is reduced to $\sigma_v^2 = 1 \deg^2$.

The same initial parameters as for the previous case are used. As verified in figure 4, the algorithm performs well under these adverse conditions. As is evident from the figure, the number of detected sources varies cleanly from $2 - 1$ and back again in the region where the tracks cross. Also, the algorithm shows no apparent tendency towards outliers in the DOA estimates in the cross region, as is commonly exhibited with other algorithms. When the sources cross, the
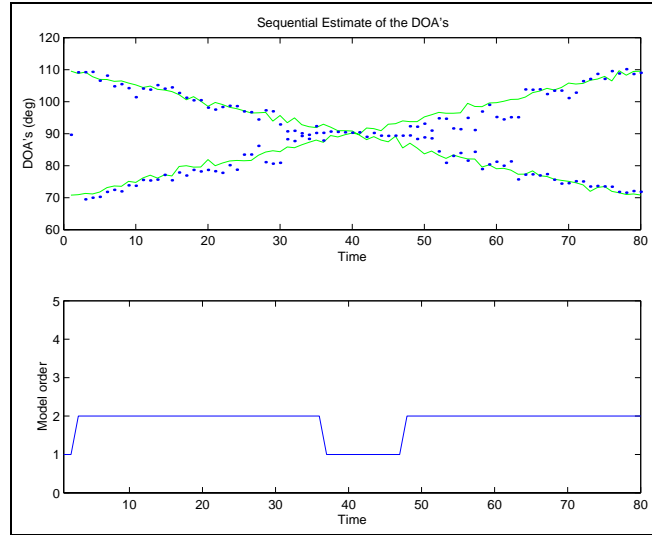
Fig. 4. Top: Sequential estimates of the directions of arrival, using the particle filter (Scenario B), and Bottom: corresponding number of detected sources.

steering matrix $\boldsymbol{S}(\cdot)$ becomes rank one and hence the two targets are seen as a single source, which explains the apparent miss-detection of a second source during the period of time when they are very close. This scenario is more difficult than the ones presented in [10], as the sources here follow steep trajectories and the variance between snapshots is high, making estimation of statistics by time-averaging very difficult. As expected under these conditions, the performance of the PAST method is significantly degraded, as shown in figure 5.

C. Data association via matching of probability distributions

The data association problem in our case becomes determining the most likely correspondence of trajectories with DOA enumeration, particularly after the merging of trajectories. In our numerical framework, we can use the estimated posterior distributions of the parameters to perform the data association task. For example, we can compare first and second moments of
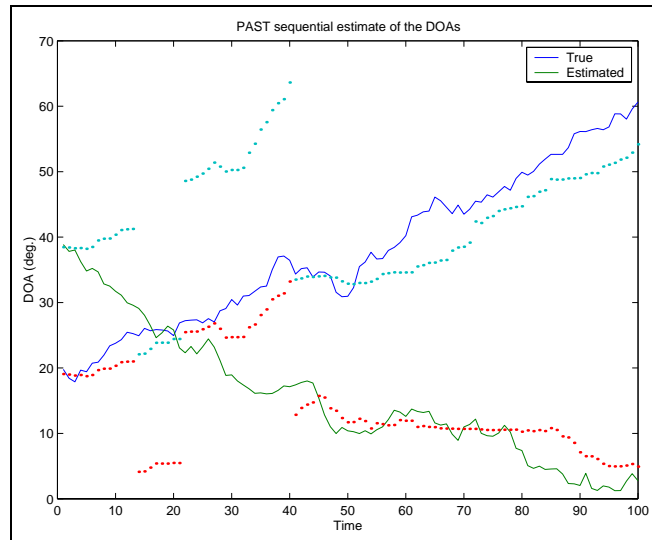
Fig. 5. Sequential estimates of directions of arrival using PAST and root-MUSIC (Scenario B).

the approximate marginal distributions of the amplitude parameters before and after crossing of the targets. Many other possibilities exist. Figure 6 shows an example of data association using the (unnormalized) marginal distribution of the amplitudes of the two sources for the above scenario of part B. The first column shows the histograms of the two amplitudes before the targets cross, while the second column shows the histograms after crossing. In this case, from visual inspection, it is clear the diagonally opposite distributions match.
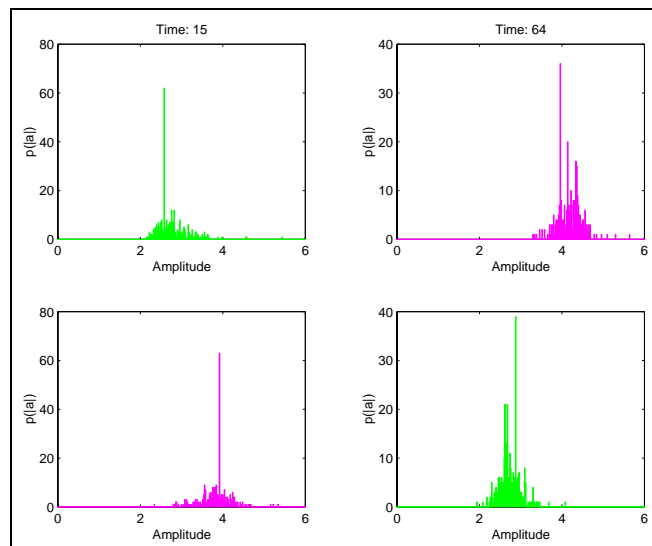

Fig. 6. Data association via the marginal posterior distributions of the amplitudes.

D. Approximate joint confidence regions

An advantage of numerical approaches to parameter estimation is that an approximation to the joint confidence region of the parameters is readily established from the histogram approximating the joint posterior distribution of the parameters. Figure 7 shows a contour plot for the joint histogram of the DOAs for the scenario of Part B above, at the 17th sample. The probability level associated with the joint confidence region is determined by integrating inside the respective contour of the normalized distribution.
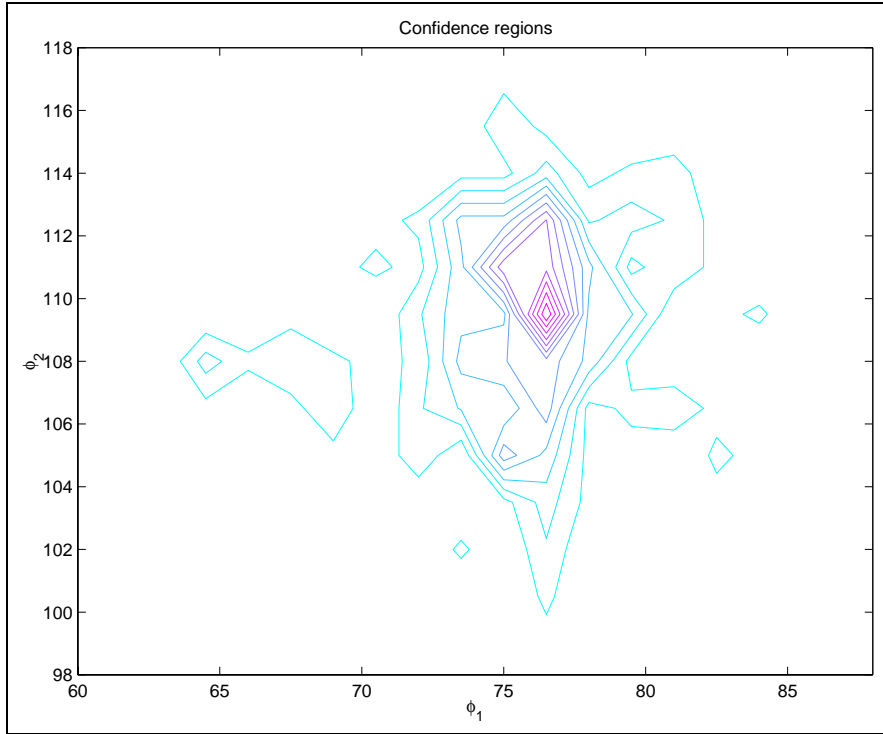

Fig. 7.   Contour plots of the approximate posterior distribution of the DOAs at the 17th sample.

E. Further Discussion

It is easily shown that as $N \to \infty$, the global optimum of the desired posterior distribution coincides with the most heavily-weighted histogram bin corresponding to the particles. In practice, the global optimum is achieved within a histogram bin-width with finite N with high probability. Thus, the global optimum can be attained by a simple search, instead of a complicated global optimization over what is shown in Figure 7 to be a multi- modal surface.

The computational expense of the particle filter approach is fairly high, requiring $\mathcal{O}(N)$ func-

tion evaluations each time step. However, the evaluation of the particles is easily parallelizable, and this order of computation does not necessarily compare unfavourably with that of a global optimization procedure. Further, the relative computational expense of the method is offset by its advantages; namely, a joint detection capability and improved performance in nonstationary environments.

VI. Conclusions

In this paper, a particle filter that includes a reversible jump MCMC with two new components, the merge and split moves, is used for joint sequential detection and estimation of an unknown number of directions of arrival.

The algorithm compares favourably to an established approach in computer simulations. The algorithm proved robust to changes in initial values and shows robust convergence to the global minima. The superior performance of the particle filter over conventional methods which use time-averaged statistics in nonstationary environments has been clearly indicated. Examples of data association of tracks before and after crossing using histogram matching, and of the joint confidence region of the parameter estimates, have been given.

References

[1] C. Andrieu, N. D. Freitas, and A. Doucet, "Sequential MCMC for Bayesian model selection," *Proceedings of HOS Conference*, 1999.

[2] B. Yang, "An extension of the PASTd algorithm to both rank and subspace tracking.," *IEEE Signal Processing Letters*, vol. 2, pp. 179–182, Sept. 1995.

[3] D. Johnson, "The application of spectral estimation to bearing estimation problem," *Proceedings of the IEEE*, vol. 70, pp. 1018–1028, Sept. 1982.

[4] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 284–280, Mar. 1986.

[6] M. Viberg and B. Ottersen, "Sensor array processing based on subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1110–1121, May 1991.

[7] J. Reilly and S. Haykin, "Maximum likelihood receiver for low-angle tracking radar Part 2: the nonsymmetric case," *IEE Proceedings*, vol. 129 Pt.F, pp. 331–340, Oct. 1982.

[8] W. Ligget, *Passive sonar: Fitting models to multiple time series.* New York Academic, 1973.

[9] Y. Zhou, P. Yip, and H. Leung, "Tracking the direction-of-arrival of multiple moving targets by passive arrays: Algorithm," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2655–2666, Oct. 1999.

[10] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.

[11] K. Molnar and J.W. Modestino, "Application of the EM algorithm for the multitarget/multisensor tracking problem", *IEEE Transactions on Signal Processing*, Vol. 46, No. 1, Jan. 1998, pp.115-129.

[12] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in practice*. Chapman and Hall, 1998.

[13] C. Andrieu, A. Doucet, W. Fitzgerald, and S. Godsill, "An introduction to the theory and applications of simulation based computational methods in Bayesian signal processing," *Tutorial ICASSP 98*, 1998.

[14] A. Doucet, "On sequential simulation-based methods for bayesian filtering.," Tech. Rep. TR.310, University of Cambridge, Signal Processing Group, 1998.

[15] C. Andrieu, P. Djuric, and A. Doucet, "Model selection by MCMC computation," *Signal Processing, Elsevier Science*, vol. 81, pp. 19–37, Jan. 2001.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[17] P. Green, "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[18] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Transactions on Signal Processing*, vol. 47, p. 2667, Oct. 1999.

[19] J. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, pp. 1032–1044, 93.

[20] J. Bernardo and A. Smith, *Bayesian Theory*. Wiley and Sons Ltd., 1994.

[21] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[22] B. Ripley, *Stochastic Simulation*. Wiley, New York, 1987.

[23] J.-R. Larocque and J. P. Reilly, "Reversible jump MCMC for joint detection and estimation of directions of arrival in coloured noise," *submitted for publications to IEEE Transactions on Signal Processing*, 2000.

[24] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[25] A. Doucet, J.F.G. de Freitas, and N.J. Gordon, Eds,. *"Sequential monte Carlo methods in practice"*, New York: Springer–Verlag, 2001.

[26] A. Doucet, N.J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems", *IEEE Trans. Signal Processing*, Vol. 49, No.3, pp. 613-624, March, 2001.