TRACKING OF MULTIPLE WIDEBAND TARGETS USING PASSIVE SENSOR ARRAYS AND PARTICLE FILTERS

Volkan Cevher, James H. McClellan

Georgia Institute of Technology
Atlanta, GA, 30332-0250
cevher@ieee.org, jim.mcclellan@ece.gatech.edu

ABSTRACT

In this paper, we present a way to track multiple maneuvering targets with varying time-frequency signatures. A particle filter is used to track targets that have constant speeds with changing heading directions. The target motion dynamics help the particle filter achieve an angular resolution otherwise not possible by the conventional beamforming techniques. Moreover, the particle filter has a built-in target association that eliminates the need for heuristic techniques commonly used in the multiple target tracking problems. Reference priors are used to derive the probability distribution function of the acoustic array outputs given the state of the multiple target states (MTS's). Local linearization is used to approximate the importance function used in the particle filter by a Gaussian pdf. Finally, computer simulations are used to demonstrate the performance of the algorithm with synthetic data.

1. INTRODUCTION

The direction-of-arrival (DOA) estimation problem has been extensively studied in the signal processing literature [1] (and references therein). Narrow-band solutions based on beamforming such as multiple signal classification (MUSIC) [2], minimum variance beamforming, and Pisarenko's method suffer in performance when the targets are moving relatively fast during the estimation batch (i.e., the snapshot period). In order to remedy this problem, one can incorporate the target motion dynamics to jointly estimate the DOA while tracking targets [3]. These refined DOA estimates provide better performance in exchange for increased complexity.

In the case of wideband acoustic signals impinging on the sensors, the pioneering work by Wang and Kaveh [4] on coherent subspace processing coherently integrates the array autocorrelation matrices corresponding to the multiple frequencies of interest, so that signal-to-noise (SNR) and resolution gains can be achieved. The work by Gershman and Amin [5] approximates the signals at the DOA batch process as chips and performs time-frequency MUSIC (spatial tf-MUSIC) on the acoustic array outputs. These wideband methods produce snapshot DOA estimates and hence require heuristics for target association.

Advances in large scale integration of computer systems have made Monte-Carlo techniques a feasible alternative as a suboptimal solution to the target tracking problem. Conventionally, given a target dynamics model, the underlying motion equations are simplified by linearization and Gaussian noise assumptions so that an analytical solution can be obtained. The extended Kalman filter is such a method; it is also the best minimum mean-squared linear estimator for the problem at hand. Monte-Carlo techniques, on the other hand, do not change the model or assume Gaussian noise; however, they approximate the posterior density of interest by particles that represent a discrete version of the posterior. The idea is that if a sufficient number of effective particles can be used, the estimation performance will be close to the theoretical optimal solution. Usually, they are suboptimal since they employ a finite number of particles.

The array model we employ in this work has a special structure. Each node of the array consists of a circular acoustic sensor array, which will supply DOA and frequency information of the targets. However, the solution is given for the general case and does not depend on the particular structure of the nodes. Spatial diversity of multiple nodes can be exploited in triangulating the targets of interest.

A solution to the problem presented thus far has been given for the narrow-band case by Orton and Fitzgerald in [6]. Our work builds on their results. Our extensions will come in the form of re-deriving the necessary gradients and Hessians used in the particle filter updates for the wideband case. It is assumed that we have a separate a time-frequency tracker for tracking the dominant instantaneous frequencies for the targets. Issues related to the uncertainty principle for frequency estimation will be left for future study.

Organization of the paper is as follows. Section 2 describes the data model used for the multiple target tracking. Sections 3 and 4 describe how we construct the probability density functions (pdf's) used by the particle filter. The importance function is discussed in section 5 and the multiple target tracking particle filter algorithm is described in section 6. Finally, section 7 shows the performance of the algorithm with synthetic data.

2. DATA MODEL

Consider K far-field targets coplanar with a sensor node consisting of P acoustic sensors. The sensor node (or sensor array) does not possess any special structure. The targets are assumed to have constant speeds with some Brownian disturbance acting on their heading directions, which is the same data model used in [3, 6].

Prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-02-0008. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

2.1. State Model

The state model has the following state vector:

$$\mathbf{x}_k(t) \triangleq \begin{bmatrix} \theta_k(t) \\ Q_k(t) \\ \phi_k(t) \\ f_k(t) \end{bmatrix} \tag{1}$$

where $\theta_k(t)$, $\phi_k(t)$, and $f_k(t)$ are the DOA, the heading direction, and the instantaneous frequency of the k^{th} target. $Q_k(t) \triangleq \log q_k(t)$ is the compound variable corresponding to the logarithmic ratio of the k^{th} target's speed (v_k) to its range ($r_k(t)$), which is measured to the center of the sensor array. Target DOA's are measured clockwise with respect to the y-axis whereas the target heading directions are measured counter clockwise with respect to the x-axis. Figure 1 illustrates the geometry of the problem.

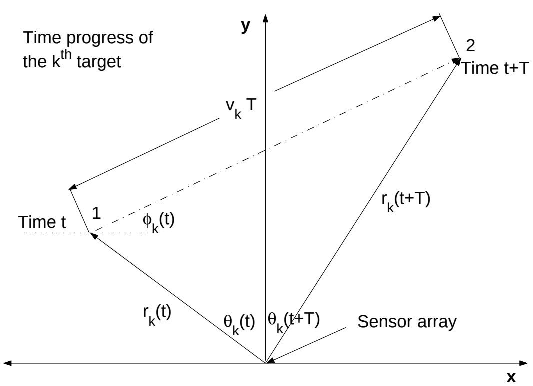

Fig. 1. The k^{th} target is at position 1 at time t and moves to position 2 in T seconds. The target is at the far-field of the sensor array whose center coincides with the origin.

The state update equation can be derived by relating the DOA's of the target at times t and $t+T$ using the geometrical relation of position 1 at $(r_k(t)\sin\theta_k(t), r_k(t)\cos\theta_k(t))$ to position 2 at $(r_k(t)\sin\theta_k(t)+v_kT\cos\phi_k(t), r_k(t)\cos\theta_k(t)+v_kT\sin\phi_k(t))$. Then, it is straightforward to obtain the following update relations:

$$\tan\theta_k(t+T) = \frac{r_k(t)\sin\theta_k(t) + v_kT\cos\phi_k(t)}{r_k(t)\cos\theta_k(t) + v_kT\sin\phi_k(t)} \tag{2}$$

and

$$r_k(t+T) = \sqrt{r_k^2(t) + 2r_k(t)v_kT\sin(\theta_k(t)+\phi_k(t)) + v_k^2T^2} \tag{3}$$

Equations (2) and (3) form a scalable system for the target motion dynamics at hand. To elaborate on this, consider scaling the range and the speed of the k^{th} target. It can be shown that this scaled target has the same set of update equations as above since the scale factor can be cancelled out. This fact, in turn, leads to the introduction of the compound variable $q_k(t) \triangleq v_k/r_k(t)$. In the state update, however, the logarithm of $q_k(t)$ is used since an additive noise component can be employed (as opposed to the multiplicative noise when $q_k(t)$ is used [1]). Hence, the state update equation can be written as follows:

$$\mathbf{x}_k(t+T) = \mathbf{f}(\mathbf{x}_k(t), \mathbf{u}(t+T))$$
$$= \begin{bmatrix} \arctan\frac{\sin\theta_k(t)+e^{Q_k(t)}T\cos\phi_k(t)}{\cos\theta_k(t)+e^{Q_k(t)}T\sin\phi_k(t)} \\ Q_k(t) - 1/2\log[1+ \\ 2e^{Q_k(t)}T\sin(\theta_k(t)+\phi_k(t)) + (e^{Q_k(t)}T)^2] \\ \phi_k(t) \\ f_k(t) + 2a_k(t)T \end{bmatrix} + \begin{bmatrix} u_{\theta,k}(t+T) \\ u_{Q,k}(t+T) \\ u_{\phi,k}(t+T) \\ u_{f,k}(t+T) \end{bmatrix} \tag{4}$$

where

$$\mathbf{u}_k \triangleq \begin{bmatrix} u_{\theta,k} \\ u_{Q,k} \\ u_{\phi,k} \\ u_{f,k} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_{\theta,k}^2 & 0 & 0 & 0 \\ 0 & \sigma_{Q,k}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\phi,k}^2 & 0 \\ 0 & 0 & 0 & \sigma_{f,k}^2 \end{bmatrix}\right) \tag{5}$$

The state noise vector is chosen to be Gaussian due to its analytical tractability (justifications can be found in [6].) Moreover, the state noise $\mathbf{u}$ is usually very small, which may lead to sample impoverishment [14] (explained in Sec. V.A.3). In fact, if the process noise is zero, the state variables can be treated as static variables in an estimation problem where using a particle filter may not be appropriate. In the discussion that follows, techniques to prevent this sample impoverishment or degeneracy will be discussed.

Implicit in (4) is a second order polynomial approximation done for the phase of the signals of interest. Hence, $2a_k(t)$ corresponds to the rate of change of the instantaneous frequency of the k^{th} signal. The rest of the paper assumes that $a_k(t)$ is supplied by a time-frequency filter and is not estimated by the particle filter.

2.2. Observation Model

The sensor array consists of P omnidirectional acoustic sensors situated uniformly on a circle of radius R. A steering vector associated with the array defines the complex array response for a source at DOA θ, and has the following form for the i^{th} source signal (the medium is assumed to be isotropic and non-dispersive):

$$\mathbf{a}(\theta_i) = \begin{bmatrix} e^{j2\pi[a_k(t)(\alpha_i{}^T\mathbf{z}_1)^2 - f_i(t)(\alpha_i{}^T\mathbf{z}_1)]} \\ e^{j2\pi[a_k(t)(\alpha_i{}^T\mathbf{z}_2)^2 - f_i(t)(\alpha_i{}^T\mathbf{z}_2)]} \\ \vdots \\ e^{j2\pi[a_k(t)(\alpha_i{}^T\mathbf{z}_P)^2 - f_i(t)(\alpha_i{}^T\mathbf{z}_P)]} \end{bmatrix} \tag{6}$$

where $i = 1, 2, \cdots, K$, $\mathbf{z}_l$ is the l^{th} sensor position, and $\alpha_i \triangleq (1/c)[\ \cos(\theta_i), \quad \sin(\theta_i)\ ]^T$ is the i^{th} slowness vector in cartesian coordinates [1]. Each steering vector $\mathbf{a}(\theta_i)$ corresponds to a signal whose direction is the objective of the DOA estimation problem. The observations are updated every $\tau = T/M$ seconds where M is the number of batch samples (a typical M is 100.) Then, the array outputs for chirp signals can be written as follows:

$$\mathbf{y}(t) = \mathbf{A}(\mathbf{\Theta}(t))\mathbf{s}(t) + \mathbf{n}(t) \qquad t = 1, 2, \cdots, M \tag{7}$$

In (7), $\mathbf{y}(t)$ is the noisy array output vector, $\mathbf{n}(t)$ is an additive noise (e.g., $\mathbf{n}(t) \sim \mathcal{N}(0, \sigma_n^2)$), and $\mathbf{A}(\mathbf{\Theta}(t))$ consists of the steering vectors in the following manner:

$$\mathbf{A}(\mathbf{\Theta}(t)) = [\ \mathbf{a}(\theta_1(t)), \quad \mathbf{a}(\theta_2(t)), \quad \cdots, \quad \mathbf{a}(\theta_K(t))\ ] \tag{8}$$

[1] due to the fact that $q_k(t)$ is a scale parameter

where $\mathbf{\Theta}(t) = [\ \theta_1(t), \quad \theta_2(t), \quad \cdots, \quad \theta_K(t)\ ]^T$. It should be noted that the sensor positions must be perfectly known in order to define $\mathbf{A}(\theta(t))$ for this model.

For notational convenience and tractability, the data collected at each time is stacked to form the following data vector $\mathbf{Y}_t$:

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t+\tau) \\ \vdots \\ \mathbf{y}(t+(M-1)\tau) \end{bmatrix} \tag{9}$$

The signal vector $\mathbf{S}_t$ and the noise vector $\mathbf{W}_t$ are formed in the same manner. Thus, the array data (or observation) model for the batch period can be compactly written as the following:

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{h}(\mathbf{\Theta}(t), \mathbf{W}_t) \\ \mathbf{Y}_t &= \mathbf{A}_t\mathbf{S}_t + \mathbf{W}_t \end{aligned} \tag{10}$$

where the steering matrix $\mathbf{A}_t = \text{diag}\{\mathbf{A}(\theta(t)), \mathbf{A}(\theta(t+\tau)), \cdots, \mathbf{A}(\theta(t+(M-1)\tau))\}$ implicity incorporates the DOA information of the targets.

3. PDF CONSTRUCTION FOR THE DATA

The particle filter is a convenient way of recursively updating a target posterior of interest. While formulating these update equations in our problem, one encounters two nuisance parameters: the signal vector $\mathbf{S}_t$ and the noise variance for the additive Gaussian noise vector $\mathbf{W}_t$. For simplicity, we will assume that the noise variance is approximately constant during the batch period $[t, t+(M-1)\tau]$. Following the notation introduced in [6] , we will denote this noise variance as $\sigma_w^2(t)$ corresponding to the batch period starting at time t. The noise has the complex Gaussian probability density function (pdf) described by Goodman [8]. The data likelihood given the signal and noise vectors can be written as follows:

$$\begin{aligned} p(\mathbf{Y}_t|\mathbf{A}_t, \mathbf{S}_t, \sigma_w^2(t)) &= \frac{1}{(\pi\sigma_w^2(t))^{MP}} \\ &\exp\Big[-\frac{(\mathbf{Y}_t - \mathbf{A}_t\mathbf{S}_t)^H(\mathbf{Y}_t - \mathbf{A}_t\mathbf{S}_t)}{\sigma_w^2(t)}\Big] \end{aligned} \tag{11}$$

If the priors are known for the signals and noise variance given the state vector at time t, they can be integrated out from the Gaussian pdf described by (11). If one desires to assume the least about these parameters and let the observed data speak for itself, then the use of reference priors comes into play[2]. Hence, even for moderate sample sizes, the information in the data dominates the *prior* information because of the vague nature of the prior knowledge [11]. One also needs to be careful about the fact that these types of priors are actually a function of the data likelihood and in general will change if, for example, new sensors are added or removed. The intuitive choice of the prior is usually the uniform prior on the *natural* space of the parameter. A good discussion of these issues can be found in [11], [12], and [10]. We will now discuss one particular case where the noise variance in (10) is known and construct the pdf's for it. The unknown variance case was treated in [6].

[2]Bernardo derives the reference prior using an estimation model based on communication channel with a source and data [9]. The reference prior maximizes the mutual information between the source and the data.

We will start by assuming that the columns of $\mathbf{A}_t$ are linearly independent. This also implies that $P > K$, i.e., the number of sensors is greater than the number of targets. Using (11), the log-likelihood function of the data can be written as follows:

$$\begin{aligned} L &\triangleq \log p(\mathbf{Y}_t|\mathbf{A}_t, \mathbf{S}_t, \sigma_w^2(t)) \\ &= -MP\log\pi - MP\log\sigma_w^2(t) \\ &\quad - \frac{1}{\sigma_w^2(t)}(\mathbf{Y}_t - \mathbf{A}_t\mathbf{S}_t)^H(\mathbf{Y}_t - \mathbf{A}_t\mathbf{S}_t) \end{aligned} \tag{12}$$

We will take the naive, easy approach in deriving the reference prior for our problem: use the square root of the determinant of the Fisher information matrix as our reference prior. After the derivations, we show that the some reference priors may not be integrable (and hence, improper) on the entire unbounded space for the parameter vectors. This in turn stipulates compactness arguments on the parameter space such as the ones used in [11, 13] . Further properties of the reference priors (or to be consistent with the literature, Jeffrey's prior) are discussed below.

The Fisher information matrix for the parameter vector $\mathbf{S}_t$ (denoted as $\mathbf{I}_L(\mathbf{S}_t)$) can be written as

$$\mathbf{I}_L(\mathbf{S}_t) = -E\Big[\frac{\partial^2 L}{\partial \mathbf{S}_t^2}\Big] = \frac{1}{\sigma_w^2(t)}\mathbf{A}_t^H\mathbf{A}_t \tag{13}$$

where $E[.]$ is the expectation operator. Hence, the reference prior in this case is

$$p(\mathbf{S}_t|\mathbf{A}_t) \propto |\mathbf{A}_t^H\mathbf{A}_t|^{1/2} \tag{14}$$

where $|.|$ is the determinant. At this point, we can use (14) to integrate out the signal vector in (11) from our problem. The integrals are from $-\infty$ to ∞ unless stated otherwise.

$$\begin{aligned} p(\mathbf{Y}_t|\mathbf{A}_t, \sigma_w^2(t)) &= \int p(\mathbf{Y}_t|\mathbf{A}_t, \mathbf{S}_t, \sigma_w^2(t))p(\mathbf{S}_t|\mathbf{A}_t)d\mathbf{S}_t \\ \propto \int \frac{1}{(\pi\sigma_w^2(t))^{MP}} &\exp\big[-\tfrac{(\mathbf{Y}_t-\mathbf{A}_t\mathbf{S}_t)^H(\mathbf{Y}_t-\mathbf{A}_t\mathbf{S}_t)}{\sigma_w^2(t)}\big]|\mathbf{A}_t^H\mathbf{A}_t|^{1/2}d\mathbf{S}_t \end{aligned} \tag{15}$$

Let $\mathbf{S}_t = (\mathbf{A}_t^H\mathbf{A}_t)^{-1}\mathbf{V}_t$. Then, $d\mathbf{S}_t = d\mathbf{V}_t/|\mathbf{A}_t^H\mathbf{A}_t|$ and (15) becomes

$$\begin{aligned} &\propto \int \frac{1}{(\pi\sigma_w^2(t))^{MP}} \\ &\exp\Big[-\frac{(\mathbf{Y}_t-\mathbf{A}_t(\mathbf{A}_t^H\mathbf{A}_t)^{-1}\mathbf{V}_t)^H(\mathbf{Y}_t-\mathbf{A}_t(\mathbf{A}_t^H\mathbf{A}_t)^{-1}\mathbf{V}_t)}{\sigma_w^2(t)}\Big]\frac{d\mathbf{V}_t}{|\mathbf{A}_t^H\mathbf{A}_t|^{1/2}} \\ &\propto \frac{1}{(\pi\sigma_w^2(t))^{M(P-K)}}\int\frac{1}{(\pi\sigma_w^2(t))^{MK}|\mathbf{A}_t^H\mathbf{A}_t|^{1/2}} \\ &\exp\Big[-\frac{(\mathbf{V}_t-\mathbf{A}^H\mathbf{Y}_t)^H(\mathbf{A}_t^H\mathbf{A}_t)^{-1}(\mathbf{V}_t-\mathbf{A}_t^H\mathbf{Y}_t)}{\sigma_w^2(t)}\Big]d\mathbf{V}_t \\ &\qquad\times\exp\Big[-\frac{\mathbf{Y}_t^H(\mathbf{I}-\mathbf{A}_t(\mathbf{A}_t^H\mathbf{A}_t)^{-1}\mathbf{A}_t^H)\mathbf{Y}_t}{\sigma_w^2(t)}\Big] \\ &\Rightarrow p(\mathbf{Y}_t|\mathbf{A}_t, \sigma_w^2(t)) \propto \exp\Big[-\frac{\mathbf{Y}_t^H(\mathbf{I}-\mathbf{A}_t(\mathbf{A}_t^H\mathbf{A}_t)^{-1}\mathbf{A}_t^H)\mathbf{Y}_t}{\sigma_w^2(t)}\Big] \end{aligned} \tag{16}$$

Finally, notice that (14) is not integrable if $\mathbf{S}_t$ has infinite multidimensional support. However, the condition $\{\mathbf{S}_t : |[\mathbf{S}_t]_i| < \gamma_i\}$ can be easily imposed on the i^{th} signal component for some large γ_i. This makes the prior (14) integrable on the signal vector space and, in turn, the marginalization integrals become approximate. This condition is always satisfied in practice (e.g., the signals of interest have finite magnitudes at all times.) Moreover, the finiteness condition on the signals implies that the array outputs are also finite (consider the discrete array model (7)). Hence, the approximate data probability distribution in (16) is proper (i.e., integrable).

4. PDF CONSTRUCTION FOR THE STATE

In the previous section, we omitted to motivate the need for constructing the pdf for the data and put the emphasis on the use of reference priors. Now, it is necessary to elaborate on the reasons for constructing the probability distributions for the data and the state. The state and observation models (4) and (10) form a hidden Markov model (HMM), which can be compactly described by the following pdf's:

$$\begin{array}{l} p(\mathbf{X}_{t+T}|\mathbf{X}_t) \\ p(\mathbf{Y}_{t+T}|\mathbf{X}_{t+T}) \end{array} \tag{17}$$

where $\mathbf{X}_t = [\mathbf{x}_1^T(t), \mathbf{x}_2^T(t), \cdots, \mathbf{x}_K^T(t)]^T$ and $p(\mathbf{Y}_{t+T}|\mathbf{X}_{t+T}) = p(\mathbf{Y}_t|\mathbf{A}_t)$ or $p(\mathbf{Y}_t|\mathbf{A}_t, \sigma_w^2(t))$ depending upon whether or not we treat the noise variance as a known parameter. Here, we introduce the common notation in the particle filtering literature, $\mathbf{z}_{0:t} \triangleq \{\mathbf{z}_0, \mathbf{z}_T, \ldots, \mathbf{z}_t\}$. The recursive update for the HMM model described by (17) can be written as follows [14]:

$$p(\mathbf{X}_{0:t+T}|\mathbf{Y}_{0:t+T}) = p(\mathbf{X}_{0:t}|\mathbf{Y}_{0:t}) \frac{p(\mathbf{Y}_{t+T}|\mathbf{X}_{t+T}) p(\mathbf{X}_{t+T}|\mathbf{X}_t)}{p(\mathbf{Y}_{t+T}|\mathbf{Y}_{0:t})} \tag{18}$$

Hence, the recursive evaluation of $p(\mathbf{X}_{0:t+T}|\mathbf{Y}_{0:t+T})$ requires the pdf's shown in (17). The previous section considered the construction of the second pdf in the model. This section will concentrate on the first pdf in (17).

The objective is to find $p(\mathbf{X}_{t+T}|\mathbf{X}_t)$ given (4). By inspection, one can see that $\mathbf{X}_{t+T}$ is also normal with mean $\mathbf{X}_t$ and covariance equal to that of the additive noise. Therefore, we can write the pdf for the state update as follows:

$$\begin{aligned} p(\mathbf{X}_{t+T}|\mathbf{X}_t) = & \frac{1}{(2\pi)^{2K}(\sigma_\theta\sigma_Q\sigma_\phi\sigma_f)^K} \\ & \exp\Big[-\frac{1}{2\sigma_\theta^2}(\mathbf{\Theta}_{t+T} - \mathbf{\Theta}_t)^2 - \frac{1}{2\sigma_Q^2}(\mathbf{Q}_{t+T} - \mathbf{Q}_t)^2 \\ & - \frac{1}{2\sigma_\phi^2}(\mathbf{\Phi}_{t+T} - \mathbf{\Phi}_t)^2 - \frac{1}{2\sigma_f^2}(\mathbf{F}_{t+T} - \mathbf{F}_t)^2 \Big] \end{aligned} \tag{19}$$

where

$$\begin{aligned} \mathbf{\Theta}_t &= [\ \theta_1(t), \quad \theta_2(t), \quad \ldots, \quad \theta_K(t)\ ]^T, \\ \mathbf{Q}_t &= [\ \mathbf{Q}_1(t), \quad \mathbf{Q}_2(t), \quad \ldots, \quad \mathbf{Q}_K(t)\ ]^T, \\ \mathbf{\Phi}_t &= [\ \phi_1(t), \quad \phi_2(t), \quad \ldots, \quad \phi_K(t)\ ]^T, \\ \mathbf{F}_t &= [\ \mathbf{f}_1(t), \quad \mathbf{f}_2(t), \quad \ldots, \quad \mathbf{f}_K(t)\ ]^T. \end{aligned} \tag{20}$$

We have two important remarks on the construction of the pdf's for our problem. The first one is that it is in general true that we need the analytical expressions for the pdf's to make use of the particle filter, which in general do not assume a Gaussian model. The second remark is on model order of the HMM. The motion equations describe a first order HMM model and hence the update equations (18) depend only on the previous state. If more complicated motion equations are formulated in the state model that increase the HMM model order, then a new recursive update formulation becomes necessary.

5. CHOICE OF THE IMPORTANCE FUNCTION

An appropriate choice of the importance function may reduce the variance of the simulation errors [3]. However, it is shown analytically ([16] and references therein) that the importance weights have increasing variance with time, which leads to increasing estimation errors (or simulation errors, will be used interchangeably). Here, we state an important result: the unconditional variance of the importance weights, i.e. with the observations $\mathbf{Y}_{0:t}$ being interpreted as random variables, increases over time. This fact is also known as the degeneracy phenomenon: after a few iterations, all but one of the normalized importance weights will be very close to zero [16].

When the notion of optimality enters into a problem, it is natural to question the optimality criterion. Any optimal solution will be as good as its objective function and, in our case, the objective is to minimize the variance of the importance weights. This, in turn, will maximize the effective number of particles at each time step, rendering the particle filter more effective given a constant number of particles.

There are many ways of approximating an optimal importance function with a suboptimal importance function and we will show one of them here. Let us denote the likelihood of the importance function as $l(\mathbf{X}_{kT}) \triangleq \log p(\mathbf{X}_{kT}|\mathbf{X}_{(k-1)T}, \mathbf{Y}_{kT})$. Assume that the usual assumptions of differentiability are satisfied by l. If we do a Taylor series expansion on this likelihood around some $\mathbf{X}$, we get the following:

$$\begin{aligned} l(\mathbf{X}_{kT}) = l(\mathbf{X}) + & \Big[\frac{\partial l(\mathbf{X}_{kT})}{\partial \mathbf{X}_{kT}}\Big]^T\Big|_{\mathbf{X}_{kT}=\mathbf{X}} (\mathbf{X}_{kT} - \mathbf{X}) \\ & + \frac{1}{2}(\mathbf{X}_{kT} - \mathbf{X})^T \Big[\frac{\partial^2 l(\mathbf{X}_{kT})}{\partial(\mathbf{X}_{kT})^2}\Big]\Big|_{\mathbf{X}_{kT}=\mathbf{X}} (\mathbf{X}_{kT} - \mathbf{X}) + h.o.t \end{aligned} \tag{21}$$

If $l''(\mathbf{X}) \triangleq \partial^2 l(\mathbf{X}_{kT})/\partial(\mathbf{X}_{kT})^2$ is negative definite[4], then, by ignoring the higher order terms in the expansion, we can fit a Gaussian curve as a suboptimal importance function. Also defining $l'(\mathbf{X}) \triangleq \partial l(\mathbf{X}_{kT})/\partial \mathbf{X}_{kT}$ for notational compactness, it is a straightforward linear algebra exercise to show that the optimal importance function can be approximated as

$$\begin{aligned} \pi(\mathbf{X}_{kT}|\mathbf{X}_{0:(k-1)T}, \mathbf{Y}_{0:kT}) &= p(\mathbf{X}_{kT}|\mathbf{X}_{(k-1)T}, \mathbf{Y}_{kT}) \\ &\approx \mathcal{N}(\mu(\mathbf{X}) + \mathbf{X}, \mathbf{\Sigma}(\mathbf{X})) \end{aligned} \tag{22}$$

where

$$\begin{aligned} \mathbf{\Sigma}(\mathbf{X}) &= -[l''(\mathbf{X})]^{-1} \\ \mu(\mathbf{X}) &= \mathbf{\Sigma}(\mathbf{X}) l'(\mathbf{X}) \end{aligned} \tag{23}$$

It is crucial to note that the optimal importance function $\pi(\mathbf{X}_{kT}|\mathbf{X}^{(i)}_{0:(k-1)T}, \mathbf{Y}_{0:kT})$ is proportional to $p(\mathbf{Y}_{kT}|\mathbf{X}_{kT}) \times p(\mathbf{X}_{kT}|\mathbf{X}^{(i)}_{(k-1)T})$ with the proportionality independent of $\mathbf{X}_{kT}$ (an observation first noted in [6]). We have previously derived

[3] i.e., if we choose the exact posterior as the importance function then the due to the nature of the data generating process the variance of the estimator is inversely proportional to the number of particles N [15].

[4] One case where $l(\mathbf{X})$ is concave this statement holds; however, this is in general not true. [3] shows ways to approximate its construction for our problem so that the negative definiteness holds.

the analytical relations for $p(\mathbf{Y}_{kT}|\mathbf{X}_{kT})$ and $p(\mathbf{X}_{kT}|\mathbf{X}_{(k-1)T})$. Moreover, define

$$\begin{aligned} l_y(\mathbf{X}_{kT}) &\triangleq \log p(\mathbf{Y}_{kT}|\mathbf{X}_{kT}) \\ l_x(\mathbf{X}_{kT}) &\triangleq \log p(\mathbf{X}_{kT}|\mathbf{X}_{(k-1)T}) \end{aligned} \quad (24)$$

then, (23) can be rewritten with new parameter set (24) as follows

$$\begin{aligned} \mathbf{\Sigma}(\mathbf{X}) &= -[l''_x(\mathbf{X}) + l''_y(\mathbf{X})]^{-1} \\ \mu(\mathbf{X}) &= \mathbf{\Sigma}(\mathbf{X})[l'_x(\mathbf{X}) + l'_y(\mathbf{X})] \end{aligned} \quad (25)$$

and the new suboptimal importance function is given below:

$$\pi(\mathbf{X}_{kT}|\mathbf{X}_{0:(k-1)T}, \mathbf{Y}_{0:kT}) \approx \mathcal{N}(\mu(\mathbf{X}) + \mathbf{X}, \mathbf{\Sigma}(\mathbf{X})) \quad (26)$$

$\mathbf{X}$ is judiciously chosen to be the mode of $p(\mathbf{X}_{kT}|\mathbf{X}_{(k-1)T}, \mathbf{Y}_{kT})$ so that $\mu(\mathbf{X}) \approx \mathbf{0}$ [16].

6. ALGORITHM DETAILS

In this section, we will give the details of our modifications to the independent partition particle filtering algorithm by Orton. The outline of Orton's algorithm is given in [6] and hence will not be repeated here. The target association problem is solved by the independence assumption on the MTS partitions. However, a minor clarification of the implementation in [6] is needed. When the necessary Hessians are calculated for the whole particle, only the pertinent portions of the Hessians are used while generating the new partition in the particle. Hence, the off-diagonal matrices in the particle Hessian corresponding to the cross partitions are ignored. After the particle is formed, the discrepancies generated by this method are augmented by the weights, which are calculated using the the full Hessians generated from the particle.

One modification is the use of the state transition probability (19) for the weighted resampling functions $q_k(\mathbf{x})$. This choice alone seems to constrain the particles by the state update equation and hence is expected to have poor performance for the maneuvering targets. However, this choice of the weighted resampling function makes sure that the created particles form a cloud around the expected mode of the target state. The maneuvering target cases, on the other hand, are handled by the absolutely critical MCMC resampling step. In the test cases we have run, the algorithm seems to better handle the maneuvers as the number of iterations increase in the MCMC resampling algorithm outlined in [6]. When the targets maneuver, the expected mode of the next state predicted by the state update equation (4) changes. At the resampling state, the particles that are closer to this changed mean survive while the particles around the predicted mean diminish. Hence, the resampling step, in effect, not only makes the particles span most of the state space, but also compensates for the effects of the maneuver. It should be noted that maneuvering has more impact on the heading direction than the other state variables. Hence, a slight modification exploiting this fact in the resampling step may also improve the performance of the algorithm for a given number of particles.

Because of the state vector is larger, the most important extension comes in the form of deriving new gradients and Hessians (25) for the linearization of the optimal importance function. The new gradients and Hessians related to the state update are straightforward and hence, we will concentrate on l'_y and l''_y. We will show the modifications to G and H following the notation in [3]. First note that[5] $l_y = -NJ/\sigma_w^2$ where J is as defined at Equation (25) in [3]. ∇J_t, $A_m(t)$, and $D(t)$ remain the same. Definition of $V(t)$ does not change; however, the vector changes due to the fact that the gradient of $\theta_k(t)$ with respect to $\alpha_k(1)$ includes an extra term $\partial\theta_k/\partial f_k(1)$. Define $\psi_k(t)$ as the gradient of $f_k(t)$ with respect to $\alpha_k(1)$, and form $\Psi(t)$ in the same manner as $V(t)$ using $\psi_k(t)$. Moreover, denote $\nabla J_f = \begin{bmatrix} \partial J_t/\partial f_1(t), & \partial J_t/\partial f_2(t), & \cdots, & \partial J_t/\partial f_K(t) \end{bmatrix}$, then the new gradient G can be written as

$$G = \text{vec}\left\{\frac{1}{N}\sum_{t=1}^{N} V(t)\text{diag}(\nabla J_t) + \Psi(t)\text{diag}(\nabla J_f)\right\} \quad (27)$$

Here we define $C_m(t) = \partial A(t)/\partial f_m$ and the approximate the new derivatives due to our state vector as follows:

$$\frac{\partial^2 J_t}{\partial\theta_m(t)\partial f_n(t)} \simeq 2\text{Re}\{\text{tr}[A^{\dagger H}(t)A_m^H(t)P_{A(t)}^{\perp}C_n^H(t)A^{\dagger H}(t)]\}$$

and

$$\frac{\partial^2 J_t}{\partial f_m(t)\partial f_n(t)} \simeq 2\text{Re}\{\text{tr}[A^{\dagger H}(t)C_m^H(t)P_{A(t)}^{\perp}C_n^H(t)A^{\dagger H}(t)]\} \quad (28)$$

Without these approximations, the Hessian matrices, which basically approximate the covariance matrices for the suboptimal importance function are not guaranteed to be positive semi-definite. The necessity of these approximations are further discussed in [3].

7. SIMULATION RESULTS

We will present a case where the algorithm performed poorly among all the simulations we run. The sensor node consists of a 15-element uniform circular array whose inter-element spacing is 0.45 λ_{min}. Other pertinent parameters are $T = 1s$, $M = 50$, $SNR = 10dB$, $N = 200$, number of MCMC steps is 5, $\sigma_\theta^2 = (.1^\circ)^2$, $\sigma_Q^2 = (.01)^2$, $\sigma_\phi^2 = (4^\circ)^2$, and $\sigma_f^2 = (.001)^2$.

Figure 2 shows the evolution of the target instantaneous frequencies, which go into a 90° turn at the beginning of the 25^{th} batch period. The dashed lines in Fig. 3 are the true target DOAs where the solid lines show the particle filter estimates. The tracking is very good until the targets maneuver; however, the particle filter still does a great job given the fact that it does not know the target signals. We found that the estimates get better as N increases due to asymptotic properties of the particle filter. It should be noted that the Rayleigh resolution at these frequencies is worse than 20°; when we used the common beamformers (MUSIC, minimum-variance, linear-predictive, and conventional) on the same synthetic data, they were not able to produce any good DOA estimates at any batch period.

8. CONCLUSIONS

We presented a way to track multiple targets with varying frequency signatures. The solution is intuitively simple: it is based on a state/observation equations couple. Moreover, the approach is general: it is shown that the state vector in [3] and [6] can be

[5] N is defined to be the batch size in [3] corresponding to M

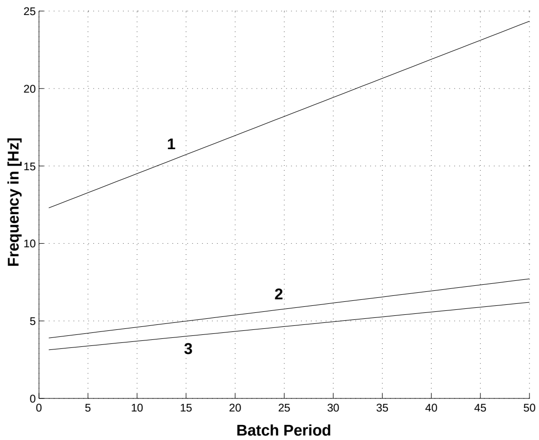

Fig. 2. The instantaneous frequency of the targets double linearly in 50 seconds. Targets 2 and 3 have very close instantaneous frequencies making them harder to track.

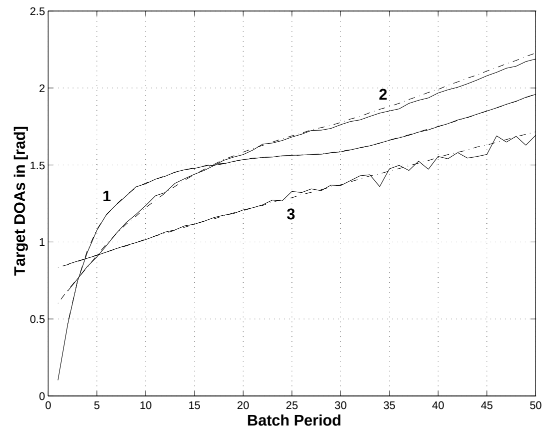

Fig. 3. The filter correctly associated the DOAs with the targets around batches 3,4,5, and 17. After the targets maneuver at the 25^{th} batch period, the DOA estimates has some bias and ripples due to the closeness of their instantaneous frequencies. $N = 200$ is used for this figure; however, the estimation performance gets better as the number of particles is increased.

extended to include other features such as the instantaneous frequency of the target signals. It is also seen that the particle filter can achieve better than Rayleigh resolution by exploiting the target motion dynamics and the previous state information of the targets. Finally, the independent partition assumption automatically takes care of the target association problem and the filter can also exploit the frequency information to separate closely moving targets.

9. REFERENCES

[1] D.H. Johnson and D.E. Dudgeon, *Array Signal Processing: Concepts and Techniques,* Prentice Hall Signal Processing Series, 1993

[2] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, May 1989, pp. 720-741

[3] Y. Zhou, P.C. Yip, H. Leung, "Tracking the direction-of-arrival of multiple moving targets by passive arrays: Algorithm," *IEEE Trans. on Signal Processing*, vol. 47, no. 10, October 1999, pp. 2655-2666

[4] H. Wang and M. Kaveh, "On the performance of signal-subspace processing-Part I:Narrow-band systems," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, October 1986, pp. 1201-1209

[5] A.B. Gershman and M.G. Amin, "Wideband direction-of-arrival estimation of multiple chirp signals using spatial time-frequency distributions," *IEEE Signal Processing Letters*, vol. 7, no. 6, June 2000, pp. 152-155

[6] M. Orton and W. Fitzgerald, "A Bayesian approach to tracking multiple targets using sensor arrays and particle filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 50, no. 2, February 2002, pp. 216-223

[7] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 50, no. 2, February 2002, pp. 174-188

[8] N.R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Annals of Mathematical Statistics*, March 1963, pp. 152-177

[9] J.M. Bernardo, "Reference posterior distributions for Bayesian inference," *J.R. Statist. Soc. B*, **41**(2), 1979, pp. 113-147

[10] J.O. Berger and J.M. Bernardo, "On the development of reference priors," *Bayesian Statistics*, **4**, 1992, pp. 35-60

[11] J.M. Bernardo and J.M. Ramon, "An introduction to Bayesian reference analysis: Inference on the ratio of multinomial parameters," *Bayesian Statistics*, **47**, 1998, pp. 101-135

[12] A.D. Lanterman, "Schwarz, Wallace, and Rissanen: Intertwining themes in theories of model selection" *International Statistical Review*, **69**(2), 2001, pp. 185-212

[13] J.J. Rissanen "Fisher information and stochastic complexity," *IEEE Transactions on Information Theory*, vol. 42, no. 1, January 1996, pp. 40-47

[14] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, February 2002, pp. 174-188

[15] A.D. Ackerberg, "A new use of importance sampling to reduce computational burden in simulation estimation," under revision at *Review of Economic Studies*, Available http://www.econ.ucla.edu/ackerber/smoonber2.pdf

[16] A. Doucet, "On sequential simulation-based methods for Bayesian filtering," *Technical report CUED/F-INFENG/TR.310*, Department of Engineering, University of Cambridge, 1998