A Survey on Joint Tracking Using Expectation-Maximization Based Techniques

Hua Lan^{a,b}, Xuezhi Wang^c, Quan Pan^{a,b}, Feng Yang^{a,b}, Zengfu Wang^{a,b}, Yan Liang^{a,b}

^a*School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, P.R.China*

^b*Key Laboratory of Information Fusion Technology, Ministry of Education, Xi'an, Shaanxi, 710072, P.R.China*

^c*School of Electrical & Computer Engineering, RMIT University, Melbourne 3000, Australia*

Abstract

Many target tracking problems can actually be cast as joint tracking problems where the underlying target state may only be observed via the relationship with a latent variable. In the presence of uncertainties in both observations and latent variable, which encapsulates the target tracking into a variational problem, the expectation-maximization (EM) method provides an iterative procedure under Bayesian inference framework to estimate the state of target in the process which minimises the latent variable uncertainty. In this paper, we treat the joint tracking problem using a united framework under the EM method and provide a comprehensive overview of various EM approaches in joint tracking context from their necessity, benefits, and challenging viewpoints. Some examples on the EM application idea are presented. In addition, future research directions and open issues for using EM method in the joint tracking are given.

Keywords: Target tracking, information fusion, joint identification and estimation, expectation maximization, Bayesian inference

1. Introduction

Target tracking is a process of information processing that provides the trajectories of one or more moving objects based on sensor data and prior information of target state. It is essential for many applications in the areas of defense, medical science, traffic control and navigation, etc [1, 2]. Based on the number of sensors involved, target tracking may be divided into single sensor or multi-sensor tracking problems, in the latter case, centralized or distributed structures are taken into account. In each case, tracking is further categorized into single target and multi-target tracking problems, where in addition to measurement noise, the uncertainties induced by false alarms, low detection probability and data association between target and measurement, etc. must be addressed. While fully covered by the theories of estimation and optimization, the advanced target tracking techniques which reflect the cutting-edge research effort are multi-disciplinary among signal processing, information theory, statistical estimation and inference, artificial intelligence, and often required to deal with various uncertainties simultaneously which obscure the underlying target states from data, such as those arise in unknown disturbance input, multi-mode, nonlinear and high dimension issues:

Email address: lanhua@nwpu.edu.cn (Hua Lan)

Preprint submitted to Information Fusion

November 15, 2015

- *Unknown disturbance input* (continuous-valued uncertainty) includes the systematic modelling errors that are represented as unknown inputs to the nominal model, such as target maneuvering, sensor system bias.
- *Multi-mode* (discrete-valued uncertainty) is used to model the multiple choice behaviors of a tracking system, which is complicated as it involves exhaustive hypothesis listing with incomplete information from data. It includes the measurement-to-target association (data association) in a clutter or multi-target tracking scenario, measurement-to-mode association for multi-detection systems, and track-to-track association for multi-sensor fusion.
- *Nonlinear* problem is very common in practical target tracking system. However, estimation involving nonlinear systems/observations is in general non-trivial as far as optimality is concerned. The optimal Bayesian solution requires the propagation of full probability density function that may incorporate undesirable system features such as multimodality, asymmetries, discontinuities and is only approached by approximation.
- *High-dimension* problem arises from the process of multisensor data fusion that combines data from multiple sensors and related information to achieve more specific inferences of multiple targets. It increases the complexity of systems. For example, the network-centric warfare (NCW) [3] consists of multiple platforms, multiple sensors for tracking multiple targets in clutter environment.

The aforementioned issues, which must be resolved while tracking, were addressed in a joint tracking framework by many researchers in literature. The interacting multiple model (IMM) algorithm [4] is popular for maneuvering target tracking where target dynamics is approximated by a jump Markov system. Advanced approaches, such as the unscented Kalman filter (UKF) [5], particle filters (PF) [6], and Gaussian sum filters (GSF) [7] are adopted to deal with nonlinear uncertainties problems. The multiple hypothesis tracker (HMT) [8, 9], probabilistic multi-hypothesis tracker (PHMT) [10], joint probabilistic data association (JPDA) [11], MCMC-based data association [12, 13] are proposed to deal with the data association uncertainty in multiple target tracking. Moreover, the finite set statistics (FISST) based probability hypothesis density (PHD) filter [14] estimates the intensity of objects and thus avoids explicit data association, but does not model full trajectories, although labelled-aided data association scheme exists. Comprehensive surveys on the maneuvering target tracking, data association, nonlinear filter can be found in [15, 16, 17, 18, 19, 20]. These algorithms, or some combination of them, are developed for specific applications. For instance, the joint probabilistic data association technique is used in the IMM extended Kalman filters resulting in a JPDA-IMM-EKF tracker for tracking of multiple maneuvering targets in radar applications. Typically, joint tracking applications contain a random structure between the state to be estimated and measurement. The structure is in general not directly observable through measurement, for example, the data-to-track association, the association between a set of maneuver modes and multiple targets in multi-target tracking, etc. The expectation-maximization (EM) [21] method provides a way to address the joint tracking problems with a naturally missing random information structure.

Roughly speaking, joint target tracking concerns two major questions: *what* and *where* the underlying target is. The former is related to target identity and the latter corresponds to target kinematic state. In target tracking context, the estimation of target kinematic state is often referred to as parameter estimation, while target identification is referred to as resolving those unknown quantities including system modes, data/track association which are associated to the

target state under multiple hypothesis. For instance, in the case of tracking a maneuvering target in clutter, target state consisting of position and velocity is the parameter to be estimated, while the identification signifies the measurement-to-track association and target maneuver mode. In general, identification and estimation are jointly considered in multi-target tracking scenarios and seen as the relationship of “chicken and egg”. As illustrated in Fig. 1, a poor identification of data association can deteriorate state estimation while the error yielded in the parameter estimation may result in a wrong target identification/labeling. A joint consideration of identification and estimation is more promising than a separate consideration [22, 23]. In the optimization theoretic parlance, this joint approach has the potential of arriving at a globally optimal solution, and its solution can always be obtained by an iterative algorithm[24]. Consequently, the formulation of the joint identification and estimation problem, which allows this problem to be solved iteratively, is crucial. In our recent work [25, 26, 27, 28, 29], target tracking is formulated as a joint tracking problem, where the underlying target state can only be observed via their relationship with a latent variable. The object of target tracking is to estimate the target state using observation conditioned on the latent variable. The state and latent variable correspond to the estimation and identification of the target, respectively. We addressed such a problem using the EM approach.

Figure 1: Joint target tracking illustration: the joint target tracking includes the recursive state estimation and mode state identification, these two components are dependent, or only conditionally independent.

The EM algorithm, formally proposed by Dempster *et. al.* in [21], is an iterative inference method. Indeed, it is an off-line, iterative, maximum likelihood method that guarantees to converge to a local maximum of the observed data likelihood function under some mild conditions. Implementation of EM algorithm is based on the observation that the maximization of complete data likelihood is generally easier than that of observed data likelihood function. The latent variables are unknown, however, the most recent knowledge is given by the posterior distribution of latent variables. Because the posterior of latent variables depends on the parameter of target state and the maximization over complete data likelihood requires the latent variables, this is indeed a “chicken-and-egg” problem. The EM algorithm uses an iterative procedure to circumvent this problem: after making an initial guess on the underlying target state, an E-step and a M-step

are alternatively operated. Both target state and latent variables are estimated and identified respectively in the iteration loops. While both the E-step and M-step can still remain intractable when the dimension of latent variables is too high [30]. This is a serious drawback since the EM cannot be applied to complex Bayesian models. The variational methodology is an iterative approach that is gaining popularity to ameliorate this shortcoming of the EM algorithm. This methodology allows inference in the case of complex graphical models, which in certain case provides significant improvements as compared to the EM algorithm [30].

In this paper, we consider the target tracking problem as a joint estimation and identification problem where the underlying target state may only be observed via the relationship with a latent variable. We show that the problem can be formulated and solved under a unified Bayesian estimation framework, though it is analytically intractable since the problem involves the *unknown disturbance input, multi-mode, nonlinear and high-dimension*. The EM based Bayesian inference method is described in the joint tracking context which provides an iterative procedure to deal with latent variable uncertainty while tracking. The EM algorithm for joint estimation and identification problem is presented in a way which hopefully can provide some insights into the underlying joint tracking problem. Several examples on joint estimation and identification applications are presented and the further works are suggested in the conclusion. For reader convenience, all notations used in this paper are summarised in Table 1.

Our contributions in this paper are highlighted as follows. (1) Comprehensive overview of various EM techniques with applications in literature, particularly, in joint tracking context; (2) We specifically formulate the joint tracking problem in a united framework under the EM method, and present two EM application examples in an attempt to give insights of the EM method for handling the problem of joint identification and estimation; (3) Discussions on open issues, ongoing research topics and outlooks of the EM-based target tracking research.

2. Bayesian Joint Tracking Problem

A multi-sensor multi-target tracking system, which consists of t targets and s sensors, is described by the following equations:

$$x_k^i = f_{k-1}^{i,r}(x_{k-1}^i, a_{k-1}^i) + \Gamma_{k-1}^i v_{k-1}^i, \quad i = 1, \dots, t. \quad (1)$$

$$y_k^{j,\mu} = h_k^{j,\tau}(x_k^j, b_k^{j,\tau}) + w_k^{j,\tau}, \quad j = 1, \dots, s. \quad (2)$$

where x_k and y_k represent the system state and measurement, and f_k , h_k , Γ_k represent state transition function, measurement function, and control matrix, respectively. The continued-value a_k and b_k are unknown disturbed inputs in dynamic model and measurement model, respectively. The process noise v_k and measurement noise w_k are zero-mean white Gaussian noises with corresponding covariance Q_k and $R_k > 0$. The initial state x_0 is Gaussian distributed with mean $\bar{x}_0$ and associated covariance Σ_0 . Here, we drop the superscript i, j, τ, μ for brevity.

Let Θ_k represents all model dependent parameters in equations (1) and (2). For given measurements $\mathcal{Y}_k \triangleq \{Y_k^1, \dots, Y_k^s\}$ with $Y_k^j \triangleq \{y_k^{j,1}, \dots, y_k^{j,m_j}\}$ on the target state $\mathcal{X}_k \triangleq \{x_k^1, \dots, x_k^t\}$, the joint target tracking problem under Bayesian inference is to find the posterior distribution $p(\mathcal{X}_k | \mathcal{Y}_k)$ [31]. In general, this problem is computationally intractable as information on Θ_k is required [32]. However, as described by John W. Tukey, "An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem". Assuming Θ_k is known exactly, given the measurement $\mathcal{Y}_k$, the underlying joint tracking problem is simply

Table 1: Notations used in this paper

$i \in \{1, 2, \dots, t\}$	index of target
$r \in \{1, 2, \dots, l_i\}$	index of motion mode
$j \in \{1, 2, \dots, s\}$	index of sensor
$\mu \in \{1, 2, \dots, m_j\}$	index of measurement
$\tau \in \{1, 2, \dots, \iota\}$	index of propagation mode
t	number of target
l_i	number of motion mode for the i th target
s	number of sensor
m_j	number of measurement from the j th sensor
ι	number of propagation mode
$\alpha_k \triangleq \{i, j\}$	the i th track to j th track association
$\beta_k \triangleq \{i, r\}$	the i th track to r th motion mode association
$\gamma_k \triangleq \{i, \mu\}$	the i th track to μ th measurement association
$\delta_k \triangleq \{\mu, \tau\}$	the μ th measurement to τ th propagation mode association
x_k^i	the i th target state
$\mathcal{X}_k \triangleq \{x_k^1, \dots, x_k^t\}$	states of all targets
$Y_k^j \triangleq \{y_k^{j,1}, \dots, y_k^{j,m_j}\}$	the j th sensor measurement set
$\mathcal{Y}_k \triangleq \{Y_k^1, \dots, Y_k^s\}$	measurements of all sensors
$\vartheta_k \triangleq \{f_k^i, h_k^{i,\tau}, R_k^i, Q_k^i, \bar{x}_0^i, \Sigma_0^i\}$	system model parameters
$\theta_k \triangleq \{\alpha_k, \beta_k, \gamma_k, \delta_k, a_k, b_k\}$	system external parameters
$\Theta_k \triangleq \{\vartheta_k, \theta_k\}$	all state dependent parameters
k	time step

144 to find the probability density function of $\mathcal{X}_k$ conditioned on $\mathcal{Y}_k$ and Θ_k , i.e.,

$$p(\mathcal{X}_k|\mathcal{Y}_k) = p(\mathcal{X}_k|\mathcal{Y}_k, \Theta_k) \quad (3)$$

145 which is often trivial to be solved.

When Θ_k is only partly known, i.e., the parameters of association hypothesis $\alpha_k, \beta_k, \gamma_k, \delta_k$, and the disturbance inputs a_k and b_k are measured through a partially observed hidden Markov process, the posterior density (3) is given by a hybrid, high dimensional integration:

$$\begin{aligned} p(\mathcal{X}_k|\mathcal{Y}_k) &= \int_{\Theta_k} p(\mathcal{X}_k, \Theta_k|\mathcal{Y}_k) d\Theta_k \\ &= \sum_{\alpha_k} \sum_{\beta_k} \sum_{\gamma_k} \sum_{\delta_k} \int_{a_k} \int_{b_k} p(\mathcal{X}_k, \alpha_k, \beta_k, \gamma_k, \delta_k, a_k, b_k|\mathcal{Y}_k) da_k db_k \end{aligned} \quad (4)$$

146 **Remark 2.1.** Joint target tracking is to solve the hybrid system (4), which involves the integra-
 147 tion of unknown disturbance inputs (continue variables a_k and b_k), and summation of multi-mode
 148 (discrete variables $\alpha_k, \beta_k, \gamma_k$, and δ_k) over the latent space Θ_k . Essentially, it is an NP-hard
 149 problem. In the case of continuous variables, the required integrations may not have closed-
 150 form analytical solutions, while complexity of the integrand may prohibit numerical integration.

151 For discrete variables, the marginalization involves summing over all possible configurations
 152 of the hidden variables, and there may be exponentially increased hidden states so that exact
 153 calculation is prohibitively expensive in practice.

154 **Remark 2.2.** This multisensor and multitarget tracking problem can be regarded as an incom-
 155 plete data problem. That is, in order to estimate the target state X_k from measurement $\mathcal{Y}_k$,
 156 knowledge of parameter Θ_k is needed. However, the parameter Θ_k is missing.

157 **Remark 2.3.** Equation (4) involves both state estimation of X_k and parameter identification of
 158 Θ_k . While the target state X_k is what we are interested in, the parameter Θ_k needs to be identified
 159 accurately. Since the identification risk and estimation error are correlated, we must perform the
 160 state estimation X_k and parameter identification Θ_k jointly.

161 Using an iterative optimization technique to retrieve incomplete data from measurement, the
 162 EM method can be used to solve the problem of joint estimation and identification under multiple
 163 uncertainties. This paper intends to provide a technical overview of the EM approaches to the
 164 joint identification and estimation problem in literature, and reviews some applications of target
 165 tracking based on the EM algorithm.

166 3. The EM Framework for Joint Tracking

167 3.1. Expectation Maximization Algorithm

168 The EM algorithm has been widely used in engineering and statistical literature as an itera-
 169 tive optimization procedure for computing maximum likelihood (ML) or maximum *a posteriori*
 170 (MAP) parameter estimates of *incomplete data problem* [21]. Since formally introduced by
 171 Dempster *et al* (1977) three decades ago, the EM algorithm has received tremendous attention
 172 from researchers across different disciplines. The popularity of EM lies mainly in its ability to
 173 find the ML or MAP estimation for many popular statistics models in a widely acclaimed manner
 174 of simplicity and stability [33]. A good survey on the EM algorithm can be found in [34]. The
 175 common derivation of EM algorithm is based on *Jensen's inequality*. Here, we briefly describe
 176 the EM algorithm under the criterion of Maximum Likelihood Estimation (MLE).

Consider a probabilistic model in which we collectively denote all of the observed variables
 by $\mathcal{Y}$ and all of the hidden variables by Θ . The joint distribution $p(\mathcal{Y}, \Theta | \mathcal{X})$ is governed by a set
 of parameters denoted $\mathcal{X}$. Our goal is to maximize the log likelihood function given by

$$L(\mathcal{X}) = \log p(\mathcal{Y} | \mathcal{X}) = \log \sum_{\Theta} p(\mathcal{Y}, \Theta | \mathcal{X}) \quad (5)$$

177 Assume the hidden variables Θ is discrete, although the discussion is identical if Θ comprises
 178 continuous variables or a combination of discrete and continuous variables, with summation
 179 replaced by integration as appropriate.

The EM algorithm is an iterative procedure for maximizing $L(\mathcal{X})$. Assume that after the n th
 iteration the current estimate for $\mathcal{X}$ is given by $\mathcal{X}^{(n)}$. Since the object is to maximize $L(\mathcal{X})$, it is
 desirable to compute an updated estimate $\mathcal{X}$ such that,

$$L(\mathcal{X}) > L(\mathcal{X}^{(n)}) \quad (6)$$

A key observation is that the summation over the latent variables appears inside the logarithm, which results in complicated expressions for the maximum likelihood solution. The EM algorithm can be derived using *Jensen's inequality*, which said

$$\log \sum_{i=1}^m \lambda_i x_i \geq \sum_{i=1}^m \lambda_i \log x_i \quad (7)$$

for constants $\lambda_i \geq 0$ with $\sum_{i=1}^m \lambda_i = 1$.

Using *Jensen's inequality*, the $L(\mathcal{X}) - L(\mathcal{X}^{(n)})$ can be written as follows

$$\begin{aligned} L(\mathcal{X}) - L(\mathcal{X}^{(n)}) &= \log \left(\sum_{\Theta} p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X}) \right) - \log p(\mathcal{Y}|\mathcal{X}^{(n)}) \\ &= \log \left(\sum_{\Theta} p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X}) \times \frac{p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)})}{p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)})} \right) - \log p(\mathcal{Y}|\mathcal{X}^{(n)}) \\ &= \log \left(\sum_{\Theta} p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) \times \frac{p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X})}{p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)})} \right) - \log p(\mathcal{Y}|\mathcal{X}^{(n)}) \\ &\geq \sum_{\Theta} p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) \log \left(\frac{p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X})}{p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)})} \right) - \log p(\mathcal{Y}|\mathcal{X}^{(n)}) \\ &= \sum_{\Theta} p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) \log \left(\frac{p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X})}{p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) p(\mathcal{Y}|\mathcal{X}^{(n)})} \right) \\ &\triangleq \Delta(\mathcal{X}|\mathcal{X}^{(n)}) \end{aligned} \quad (8)$$

Let $l(\mathcal{X}|\mathcal{X}^{(n)}) \triangleq L(\mathcal{X}^{(n)}) + \Delta(\mathcal{X}|\mathcal{X}^{(n)})$, then equation (8) shows explicitly $L(\mathcal{X}) \geq l(\mathcal{X}|\mathcal{X}^{(n)})$. An interpretation of EM algorithm is given by Fig. 2.

We emphasize that the object is to choose a value of $\mathcal{X}$ so that $L(\mathcal{X})$ is maximized. Directly optimizing of $L(\mathcal{X})$ is difficult, but optimizing the lower bound function $l(\mathcal{X}|\mathcal{X}^{(n)})$ is often significantly easier. Therefore, the EM algorithm calls for selecting $\mathcal{X}$ such that $l(\mathcal{X}|\mathcal{X}^{(n)})$ is maximized. That is,

$$\begin{aligned} \mathcal{X}^{(n+1)} &= \arg \max_{\mathcal{X}} l(\mathcal{X}|\mathcal{X}^{(n)}) \\ &= \arg \max_{\mathcal{X}} \left\{ L(\mathcal{X}^{(n)}) + \sum_{\Theta} p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) \log \frac{p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X})}{p(\mathcal{Y}|\mathcal{X}^{(n)}) p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)})} \right\} \\ &= \arg \max_{\mathcal{X}} \left\{ \sum_{\Theta} p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) \log p(\mathcal{Y}|\Theta, \mathcal{X}) p(\Theta|\mathcal{X}) \right\} \\ &= \arg \max_{\mathcal{X}} \left\{ \sum_{\Theta} p(\Theta|\mathcal{Y}, \mathcal{X}^{(n)}) \log p(\mathcal{Y}, \Theta|\mathcal{X}) \right\} \\ &= \arg \max_{\mathcal{X}} \mathbb{E}_{\Theta|\mathcal{Y}, \mathcal{X}^{(n)}} (\log p(\mathcal{Y}, \Theta|\mathcal{X})) \end{aligned} \quad (9)$$

The conditional expectation $\mathbb{E}_{\Theta|\mathcal{Y}, \mathcal{X}^{(n)}} (\log p(\mathcal{Y}, \Theta|\mathcal{X}))$ is also called Q function written as $Q(\mathcal{X}|\mathcal{X}^{(n)})$, which takes the expectation of complete-data likelihood $\log p(\mathcal{Y}, \Theta|\mathcal{X})$ with respect

Figure 2: Graphical interpretation of the iteration process of EM algorithm: the EM algorithm computes a lower bound on the log likelihood using the values of current parameter and then updates the new values of parameter by maximizing this bound. This process may iteratively repeat until the difference between “current” and “updated” values of parameter is sufficiently small.

186 to the latent variables Θ . The EM algorithm thus completes the MLE by iteratively operating the
 187 following two steps:

- 188 • *E-step*: Determine the Q-function $Q(X|X^{(n)})$;
- 189 • *M-step*: Maximize the Q-function $Q(X|X^{(n)})$ with respect to X .

190 **Remark 3.1.** More generally, a prior $p(X)$ may be incorporated so that the EM will maximize a
 191 posteriori density rather than likelihood:

$$X^{MAP} = \arg \max_X \log p(X|\mathcal{Y})p(X) = \arg \max_X \log p(X, \mathcal{Y}) \quad (10)$$

192 Therefore, the EM performs MAP estimation by modifying the E-step:

- 193 • *E-step*: Replace the complete-data likelihood $l(X|X^{(n)})$ with joint probability density func-
 194 tion $p(X, \mathcal{Y}, \Theta)$, and evaluate the conditional expectation of $p(X, \mathcal{Y}, \Theta)$ with respect to Θ to
 195 determine the Q-function $Q(X|X^{(n)})$;

Remark 3.2. The two steps of EM can also be derived using the variational Bayesian method. In the view of variational Bayesian inference, the log-likelihood in (5) can be written as

$$\log p(\mathcal{Y}|X) = F(q, X) + KL(q||p) \quad (11)$$

with

$$F(q, X) = \sum_{\Theta} q(\Theta) \log \left(\frac{p(\mathcal{Y}, \Theta|X)}{q(\Theta)} \right) \quad (12)$$

and

$$KL(q||p) = - \sum_{\Theta} q(\Theta) \log \left(\frac{p(\Theta|X, Y)}{q(\Theta)} \right) \quad (13)$$

where $q(\Theta)$ is any probability density function, $KL(q||p)$ is the Kullback-Leibler (KL) divergence between $p(\Theta|X, Y)$ and $q(\Theta)$. Note that the Kullback-Leibler divergence is asymmetric, non-negative and thus $\log p(Y|X) \geq F(q, X)$. In other words, $F(q, X)$ is the lower bound of the log-likelihood. Equality holds only when the Kullback-Leibler divergence vanishes, which occurs when $q(\Theta) = p(\Theta|X, Y)$. The EM algorithm and some recent advances in deterministic approximations for Bayesian inference can be viewed in the light of the decomposition in (11) as the maximization of the lower bound $F(q, X)$ with respect to the density q and the parameter X . From this perspective, the EM algorithm is viewed as lower bound maximization.

In particular, the EM is a two step iterative algorithm that maximizes the lower bound $F(q, X)$ and further maximizes the log-likelihood. Assume that the current value of the parameter is X^{OLD} . In the E-step the lower bound is maximized with respect to $q(\Theta)$. It is easy to see that this happens when $KL(q||p) = 0$, that is, when $q(\Theta) = p(\Theta|X, Y)$. In this case the lower bound is equal to the log-likelihood. In the subsequent M-step, $q(\Theta)$ is held fixed and the lower bound $F(q, X)$ is maximized with respect to X to obtain some new value X^{NEW} . This will cause the lower bound to increase and as a result, the corresponding log-likelihood will also increase.

Remark 3.3. The EM algorithm involves the two iterative steps E-step and M-step, where the E-step computes the conditional expectation of hidden variables Θ given the estimated X , and the M-step estimates the X using the identified Θ in the E-step. These two iterative processing framework of EM algorithm correspond exactly to the estimation and identification problem with the target tracking. Due to the inherent iterative convergence characteristics of EM algorithm, the closed feedback loop between state estimation X and identification Θ has been established, which is desirable for dealing with the couple problem between estimation and identification.

Remark 3.4. The EM algorithm requires that $p(\Theta|X, Y)$ is explicitly known, or at least we should be able to compute the conditional expectation of its sufficient statistics Q -function. While $p(\Theta|X, Y)$ is in general much easier to infer than $p(Y|X)$, in many interesting problems, especially when the hidden variables are of high-dimensions, it is still intractable to compute the $p(\Theta|X, Y)$ and thus the EM algorithm is not applicable.

After it was formalized and generalized by Dempster *et al*, the EM algorithm has received tremendous attention in many different research areas. Great efforts were made to simplify the implementation of EM in some situations (see Section 3.1.3), in particular, when there is no analytical solution available for the calculation of $Q(X|X^{(n)})$ in the E-step or for the maximization of it in the M-step. Here, we briefly review several developments of the EM algorithm, including its convergence, initialization, extensions and properties.

3.1.1. Convergence of the EM algorithm

The appealing property of EM algorithm is that the likelihood function monotonously increases with respect to the iteration number before reaching the local optimization value. However, the monotonicity of the EM algorithm guarantees that as EM iterates, its guesses won't get worse in terms of their likelihood, but the monotonicity alone cannot guarantee the convergence of the

sequence $\{\mathcal{X}^{(n)}\}$ with n being the n th iteration. As stated in [33], there is no general convergence theorem for the EM algorithm: the convergence of the sequence $\{\mathcal{X}^{(n)}\}$ depends on the characteristics of likelihood function, Q -function, and also the initial point $\mathcal{X}^{(0)}$. Under certain regularity conditions, Dempster *et al* [21] proved that $\{\mathcal{X}^{(n)}\}$ converges to a stationary point with a linear convergence rate. The subsequent work of Wu [35] established some of the most general convergence results for the EM algorithm, which showed that if the likelihood is unimodal and certain regularity conditions hold, then the EM algorithm converges to the unique global optimum. Furthermore, if the unobserved complete-data specification can be described by a curved exponential family with compact parameter space, all the limit points of any EM sequence are stationary points of the likelihood function. See also the more recent papers [36, 37, 38, 39]. In some estimation problems with constrained parameter spaces, the parameter value maximizing the likelihood is on the boundary of the parameter space, and these posterior constraints can greatly improve the performance over standard baselines and be competitive with more complex, intractable models, and the convergence is also guaranteed if the constraints are convex [40, 41]. In addition, the convergence rate of the EM algorithm is governed by the largest eigenvalue of the matrix of fractions of missing information due to incomplete data, and the componentwise rates of convergence can differ from each other when the fractions of information loss vary across different components of a parameter vector [42].

3.1.2. Initialization of the EM algorithm

The EM algorithm is often sensitive to the choice of the initial parameter vector, efficient initialization is an important preliminary process for the future convergence of the algorithm to the best local maximum of the likelihood function [43]. All initialization strategies can be generally classified as deterministic or stochastic. Some popular deterministic initialization approaches choose starting values based on the solution obtained from hierarchical clustering, model-based Gaussian hierarchical clustering [44], or the multistage procedure based on finding the best local modes. A considerable disadvantage of all deterministic methods is their incapability to propose another starting point. The proposed starting value may lead to an incorrect solution or even no solution when the likelihood function is unbounded. Stochastic initialization strategies do not share this shortcoming as they normally allow restarting from another point of the parameter space. The general idea is to try different starting values of parameters and choose the one that yields the largest local maximum. Because of the need to repeat the initialization step several times, these procedures are typically more time consuming than deterministic approaches. Among the well-known stochastic initialization methods the emEM [45] and RndEM [46] algorithms, which share the common idea of trying different initial values of parameters and choosing the best one that yields the largest local maximum.

3.1.3. Extensions of the EM algorithm

The EM algorithm breaks down the potentially difficult problem of maximizing the likelihood function into two stages, the E-step and the M-step, each of which will often prove simpler to implement. Nevertheless, for complex models it may be the case that either the E-step or the M-step, or indeed both, remain intractable. This leads to several extensions of the EM algorithm which are briefly described below. More details can be found in [47].

The *generalized EM (GEM)* [48] algorithm addresses the problem of an intractable M-step. Instead of aiming to maximize Q -function $Q(\mathcal{X}|\mathcal{X}^{(n)})$ with respect to $\mathcal{X}$, it seeks the parameters to increase the value of Q -function. Again, because $Q(\mathcal{X}|\mathcal{X}^{(n)})$ is a lower bound on the log likelihood

function, each complete E-step and M-step cycle of the GEM algorithm is guaranteed to increase the value of the log likelihood.

The *expectation conditional maximization (ECM)* [49] algorithm replaces the complicated M-step of EM by a number of computationally simpler conditional maximization (CM)-steps because the Q-function is difficult to be optimized directly over parameter space. Since each CM-step increases the value of Q-function, the ECM is a GEM based algorithm and thus preserves the appealing convergence property – always increasing the value of likelihood.

The *expectation conditional maximization either (ECME)* [50] algorithm is an extension of ECM, which further partitions the CM-steps into two groups ψ_Q and ψ_L with $\psi_Q \cup \psi_L = \{1, \dots, S\}$. While the CM-steps indexed by $s \in \psi_Q$ (referred to as the *MQ*-steps) remain the same with ECM, the CM-steps indexed by $s \in \psi_L$ (referred to as the *ML*-steps) maximize incomplete-data likelihood function in the subspace. It has been shown theoretically and empirically that ECME typically has a greater speed of convergence than ECM, and enjoys the same stability as EM with typically higher efficiency than EM.

The *dynamic expectation conditional maximization either (DECME)* algorithm proposed in [51] extends ECME by dynamically constructing a partition of the parameter space with a lower dimensional subspace over which an *ML*-step is conducted. It adds a dynamic CM-step to compute parameter Θ by maximizing $L(\Theta|\mathcal{Y})$ over some subspaces determined by the information collected from the CM-steps and the past iterations. Dynamically constructing the subspaces for *ML*-steps of ECME is promising for accelerating the EM algorithm. The implementation of DECME is equivalent to a conjugate direction method, which provides theoretical justification for its fast convergence.

The *Monte Carlo EM (MCEM)* [52] algorithm addresses the problem of an intractable E-step. Instead of computing the intractable analytical solution of $Q(X|X^{(n)})$, the MCEM algorithm approximates it via the Monte Carlo method. The implementation of MCEM which can strike a good balance between efficiency and accuracy relies on a smart choice of the sample size in the E-step. In general, the MCEM algorithm converges almost surely to the standard EM auxiliary function thanks to the law of large numbers.

The *parameter expanded EM (PX-EM)* [53] algorithm use a “covariance adjustment” to correct the analysis of the M-step by expanding the complete-data mode $p(\mathcal{Y}, \Theta|X)$ to a larger model $p(\mathcal{Y}, \Theta|X, \kappa)$, and κ is an auxiliary parameter whose value is fixed at κ_0 in the original model. This augmentation often provides extra information that can be used for adjustment to improve the M-step estimation. The PX-EM algorithm shares the simplicity and stability of ordinary EM, but has a faster rate of convergence since its M-step performs a more efficient analysis.

The *accelerated EM (AEM)* [54] algorithm is developed by appending a line search to each EM iteration. Formally, given initial value X_0 , in the $(n+1)$ th iteration, the new estimate $X^{(n+1)}$ is computed by $X^{(n+1)} = X^{(n)} + \zeta^{(n)} d^{(n)}$, where $d^{(n)}$ is a direction composed from the current direction and the previous directions, and $\zeta^{(n)}$ is a step size typically computed from a line maximization of the complete-data likelihood. The AEM algorithm usually converges much faster than EM, since conjugate direction method is considered to be one of the best general purpose optimization methods in terms of both stability and efficiency.

The *incremental EM (IEM)* algorithm [55] partitions data into blocks under the assumption that observations are independently and identically distributed. It traverses in a cyclic manner while incrementally updating expected complete data log-likelihood and parameters after visiting each of the blocks. The IEM claims an improved convergence rate over standard EM implementation by breaking the complete data likelihood into a product of conditional densities. Moreover, the IEM is suitable for distributed estimation in sensor networks due to its incremental form.

325 The *newscast EM (NEM)* [56] is a gossip-based distributed algorithm, which operates on net-
 326 work topologies where each node observes a local quantity and can communicate with other
 327 nodes in an arbitrary point-to-point fashion. The main difference between NEM and the standard
 328 EM algorithm is that the M-step is implemented in a decentralized manner: (random) pairs of
 329 nodes repeatedly exchange their local parameter estimate and combine them by (weighted) av-
 330 eraging. Under this protocol, nodes converge exponentially fast to the correct estimate in each
 331 M-step of the EM algorithm. Differing from IEM which carries out computation node by node
 332 at any time step, all nodes in NEM are running the same protocol in parallel. This results in a
 333 batched M-step that has at most logarithmic runtime with respect to the number of nodes.

334 3.1.4. Others of the EM algorithm

335 Cappé and Moulines [57] proposed an on-line variant of the EM algorithm which make it
 336 possible to estimate the parameters of a latent data model without storing the data. In particu-
 337 lar, each iteration of the algorithm is decomposed into two steps, where the first is a stochastic
 338 approximation version of the E-step aimed at incorporating the information that is brought by
 339 the newly available observation, and the second step consists in the maximization program that
 340 appears in the M-step of the standard EM algorithm. The on-line EM algorithm is usually sim-
 341 ple and is shown to achieve convergence to the stationary points. Denœux proposed the fuzzy
 342 EM (FEM) algorithm [58] and/or evidential EM (E2M) [59] algorithm to deal with parameter
 343 estimation from a postulated parametric static model for data with large uncertainties. Jiang *et al*
 344 [60] proposed the E-MS algorithm for model selection in the presence of missing data, and
 345 proved numerical convergence of the E-MS algorithm as well as consistency in model selection
 346 of the limiting model of the E-MS convergence.

347 3.1.5. Properties of the EM algorithm

348 The EM algorithm has several appealing properties relative to other iterative algorithms such as
 349 Newton-Raphson and Fisher’s scoring method for finding ML estimation. Some of its advantages
 350 compared to its competitors are as follows [34]:

- 351 (a) Under fairly general conditions, the convergence of EM algorithm is robust to initialisation.
 352 That is, starting from an arbitrary point in the parameter space, EM can almost always con-
 353 verge to a local maximizer.
- 354 (b) The EM algorithm is generally easy to program, and requires small storage space and can
 355 generally be carried out on a small computer.
- 356 (c) Many extensions of the EM algorithm in literature deal with the complex situations where a
 357 closed form formulation does not exist. These extensions share the stable monotone conver-
 358 gence of EM algorithm.
- 359 (d) The EM algorithm is numerically stable, with each EM iteration increasing the likelihood.
 360 In fact, the performance of an EM algorithm in terms of convergence rate and programming
 361 errors is indicated by the monotone increase in likelihood over iterations.

362 Some of the criticisms of the EM algorithm are as follows:

- 363 (a) The EM iteration does not automatically provide an estimate of the covariance matrix of the
 364 parameter estimates. However, this disadvantage can be easily removed by using appropriate
 365 methodology associated with the EM algorithm.

- 366 (b) The EM algorithm may converge slowly for some nonlinear systems with large incomplete
 367 uncertainties, and may be analytically intractable, especially, when missing data is of high-
 368 dimension.
- 369 (c) The EM algorithm does not guarantee convergence to the global maximum when multiple
 370 maxima exist. While additional procedures such as simulated annealing can be used to over-
 371 come such situations, additional computational complexity may be significant.

372 3.2. The EM based Bayesian inference in the joint tracking context

It is well known that Kalman filter (KF) can build optimal recursive estimators for linear dynamic systems with Gaussian independent noises, i.e.,

$$x_k = F_{k-1}x_{k-1} + \Gamma_{k-1}v_{k-1} \quad (14)$$

$$y_k = H_kx_k + w_k \quad (15)$$

373 where x_k , v_k and w_k are all Gaussian and mutually independent.

374 Compare the multi-sensor multi-target tracking system described by equations (1) - (2) with
 375 equations (14) - (15), the difference is that the parameter Θ is partially known in multi-sensor
 376 multi-target tracking. The single sensor single target tracking without clutter is a special case
 377 where Θ is completely known. The tracking system may be implemented by a standard KF. In
 378 such a case, the KF is regarded as the complete data estimation problem. For a general tracking
 379 problem, the parameter Θ may not always be known completely. For example, the uncertainty
 380 of Θ arises in maneuvering target tracking applications, and data association for multi-target
 381 tracking in clutter, etc. Clearly, we face a typical joint tracking problem, i.e., we need to estimate
 382 both target state and Θ from measurements. Now, we describe the procedure for solving this
 383 joint tracking problem in the EM framework as follows.

384 **Step 1: Model the variables in the EM framework:** Specifically, the incomplete data is
 385 the sensor measurement $\mathcal{Y}_{1:K}$ up to time K , the estimated parameter is the target state $\mathcal{X}_{1:K}$, and
 386 the missing data is the identified unknown parameter $\Theta_{1:K}$. In this case, the EM algorithm is
 387 used to solve the state estimation in MAP sense. It is worth mentioning that different modeling
 388 of variables gives rise to different implementations of the E-step and M-step, and thus results in
 389 different algorithms which are suitable for different applications.

Step 2: Work out the complete-data log-likelihood or joint probability density function:
 In MAP sense, the density function for joint complete data $\mathcal{Y}$, Θ , and the estimated parameter $\mathcal{X}$
 can be constructed as:

$$\begin{aligned} \log p(\mathcal{X}_{1:K}, \mathcal{Y}_{1:K}, \Theta_{1:K}) &= \log p(\mathcal{Y}_{1:K} | \mathcal{X}_{1:K}, \Theta_{1:K}) p(\mathcal{X}_{1:K} | \Theta_{1:K}) p(\Theta_{1:K}) \\ &= \log \left\{ \prod_{k=1}^K p(\mathcal{Y}_k | \mathcal{X}_k, \Theta_k) \prod_{k=1}^K p(\mathcal{X}_k | \mathcal{X}_{k-1}, \Theta_k) p(\mathcal{X}_0 | \Theta_{1:K}) p(\Theta_{1:K}) \right\} \end{aligned} \quad (16)$$

390 In practical tracking scenarios, the probability density functions $p(\mathcal{Y}_k | \mathcal{X}_k, \Theta_k)$, $p(\mathcal{X}_k | \mathcal{X}_{k-1}, \Theta_k)$,
 391 and $p(\mathcal{X}_0 | \Theta_{1:K})$ are usually assumed to be Gaussian.

Step 3: (E-step) Calculate the Q-function: Take the expectation of $\log p(\mathcal{X}_{1:K}, \mathcal{Y}_{1:K}, \Theta_{1:K})$
 with respect to $\Theta_{1:K}$, we obtain the Q-function:

$$Q(\mathcal{X}_{1:K} | \mathcal{X}_{1:K}^{(n)}) = \mathbb{E}_{\Theta_{1:K} | \mathcal{X}_{1:K}^{(n)}, \mathcal{Y}_{1:K}} \log p(\mathcal{X}_{1:K}, \mathcal{Y}_{1:K}, \Theta_{1:K}) \quad (17)$$

At the n th iteration, by calculating the posterior probability of missing data, we identify the unknown parameters $\Theta_{1:K}^{(n)}$. It depends on the n th iterative state estimate $\mathcal{X}_{1:K}^{(n)}$ and their corresponding measurements $\mathcal{Y}_{1:K}$:

$$\Theta_{1:K}^{(n)} = \mathcal{F}(\mathcal{X}_{1:K}^{(n)}, \mathcal{Y}_{1:K}) \quad (18)$$

This step is nontrivial for systems with a complex mode structure because of the evaluation of conditional probability density function of missing data $p(\Theta_{1:K}|\mathcal{X}_{1:K}^{(n)}, \mathcal{Y}_{1:K})$. In particularly, when the missing data is of high-dimension.

Step 4: (M-step) Maximize the Q function: After E-step, the Q function only relates to the target state $\mathcal{X}_{1:K}$. The $(n + 1)$ th estimated target state $\mathcal{X}_{1:K}^{(n+1)}$ is obtained by maximizing the Q function with respect to target state:

$$\mathcal{X}_{1:K}^{(n+1)} = \arg \max_{\mathcal{X}_{1:K}} Q(\mathcal{X}_{1:K}|\mathcal{X}_{1:K}^{(n)}) \quad (19)$$

A Kalman or nonlinear smoother is normally used to solve (19) in linear or nonlinear case.

Step 5: Run the E-step and M-step iteratively: The E-step and M-step must run alternately in the iteration, since the equations (18) and (19) are related with each other. Starting from an initial guess on target state $\mathcal{X}_{1:K}^{(0)}$, the unknown parameter $\Theta_{1:K}^{(n)}$ is estimated in the E-step via equation (18) at the n th iteration, which is then used to estimate the $(n + 1)$ th iterative target state $\mathcal{X}_{1:K}^{(n+1)}$ in the M-step via equation (19). The E-M steps are repeated until the values of likelihood functions in two consecutive iterations are close enough or the number of iterations reaches a pre-defined upper bound.

Step 6: Report the target state $\mathcal{X}$ and unknown parameters Θ jointly.

In this paper, we have investigated many target tracking applications which use the EM technique with various types of missing data $\phi \in \Theta$. In fact, EM method has been widely used in target tracking applications, ranging from single sensor target tracking, multi-sensor target tracking, to distributed target tracking in sensor networks, etc. The type of missing data $\phi \in \Theta$ used in these applications is summarised in Table 2.

Table 2: Taxonomy of EM algorithm in target tracking

Category	Problems	Missing data ϕ	Reference
Signal sensor target tracking	Parameter estimation for linear dynamic systems	$\phi \triangleq \{F, H, R, Q\}$	[61, 62]
	Parameter estimation for nonlinear dynamic system	$\phi \triangleq \{f, h, R, Q\}$	[63, 64, 65]
	Joint estimation for jump Markov systems	$\phi \triangleq \{\beta\}$	[66, 67, 68, 69]
	Joint estimation for linear system with unknown inputs	$\phi \triangleq \{a, b\}$	[26]
	Maneuvering target tracking with data association uncertainty	$\phi \triangleq \{\beta, \gamma\}$	[70, 71, 72]
	Multiple target tracking	$\phi \triangleq \{\gamma\}$	[10, 73, 74, 75]
	Multiple detection system target tracking	$\phi \triangleq \{\gamma, \delta\}$	[25, 27, 29, 76]
Multiple sensors target tracking	Multisensor multitarget tracking	$\phi \triangleq \{\alpha, \gamma\}$	[77, 78, 79]
	Multisensor multitarget tracking with sensor bias	$\phi \triangleq \{\bar{x}_0, \Sigma, \gamma, b\}$	[80, 81, 82]
	Multisensor fusion for jump Markov nonlinear systems	$\phi \triangleq \{\bar{x}_0, \Sigma, Q, R, \beta, b\}$	[83]
Target tracking in sensor networks	Distributed EM for Gaussian Mixtures clustering	$\phi \triangleq \{\gamma\}$	[84, 85, 86, 87]
	Joint association, registration and fusion for distributed tracking	$\phi \triangleq \{\bar{x}_0, \Sigma, \gamma, b\}$	[88]
Other joint tracking	Joint identification and classification	—	[89]
	Joint tracking and localization	—	[90]
	Joint detection and tracking (TBD)	—	[91]

3.2.1. Single sensor target tracking

Parameter estimation for linear dynamic systems: It is known from the theory that the KF is optimal in case that the model perfectly matches the real system, the entering noise is white and the covariances of the noise are exactly known. That is, the parameter ϕ is completely known. Shumway and Stoffer [61] made the first attempt at an EM algorithm for smoothing and forecasting in linear dynamical systems where the parameter ϕ is partly known (only the measurement matrix H is known), this work was further modified by Ghahramani and Hinton [62], which presented a basic form of the EM algorithm with the H unknown.

Parameter estimation for nonlinear dynamic systems: Consider the nonlinear state estimation problem with possibly non-Gaussian process noise in the presence of a certain class of measurement model uncertainty, Amin and Thia [63] proposed an EM-PF algorithm that casts the problem in a joint state estimation and model parameter identification framework. The E-step is implemented by a particle filter that is initialized by a Monte Carlo Markov chain algorithm, while the M-step estimates the parameters of the mixture of Gaussian, which is used to approximate the nonlinear observation equation. The EM-PF is used to solve a highly nonlinear bearing-only tracking problem and multi-sensor joint registration problem. Karimi *et al* [64] proposed fully-Laplace approximation EM (FLAEM) algorithm for simultaneous estimation of model parameters, process disturbance intensities, and measurement noise variances for nonlinear dynamic systems that are described by stochastic differential equations. The time-varying tracking model under the nonlinear state-space evolved system was addressed by Lei *et al* [65], where the EM algorithm is used to identify the state transition matrix F and the process noise covariance Q online by assumption that the state evolution and the likelihood of measurements data can be represented by Gaussian distribution.

Joint estimation for jump Markov systems: Consider the problem of joint state estimation and identification for jump Markov nonlinear system, Özkan *et al* [66] proposed a new Rao-Blackwellized particle filter to exploit the inherent structure of jump Markov nonlinear system. Together with the online EM algorithm, this algorithm enables recursive identification of unknown model parameters including the transition probability matrix. A jump Markov linear system was considered by Logothetis *et al* [67] and three EM-types algorithms were proposed. The first EM algorithm yields the MAP estimate for the entire sequence of the finite state Markov chain. The Second EM algorithm yields the MAP estimate for the state of a jump linear system. The third EM algorithm computes the joint MAP estimates for both discrete and continuous states. Proofs of convergence of these algorithm were given. Motivated by maneuvering target tracking, Johnston and Krishnamurthy [68] proposed an ECM based re-weighted IMM (RIMM) algorithm to modify the posterior probability of each mode, and Pulford *et al* [69] proposed two EM-type algorithms where the manoeuvres of a target is modeled as unknown input (acceleration) terms entering linearly into the system and Markov state sequence, respectively.

Stochastic dynamic systems with unknown inputs : Consider the joint estimation and identification problem of a class of discrete-time stochastic systems with unknown inputs in both the plant and sensors, Lan *et al* [26] proposed an EM based iterative optimization method. The system state is estimated by using the KS in E-step, and the analytical solution of unknown inputs is obtained in the M-step by maximizing the Q -function. The proposed method is used to solve the maneuvering target tracking in electronic counter environments.

Maneuvering target tracking with data association uncertainty : Logothetis *et al* [70] considered a single maneuvering target tracking in clutter, where both uncertain origin of the measurements and the maneuvering command are uncertainty. The proposed scheme combines a hidden

Markov model smoother (HMMs) and a Kalman smoother (KS), whereby the E-step computes the joint posterior probability density of association and control input by using HMMs, and the M-step obtains the MAP sequence of the target state based on a modified state-space model via the KS. Tracking of multiple maneuvering targets was addressed by Ruan and Willett [71] using the PMHT algorithm with multiple modes, and its performance in terms of tracking accuracy and computational load lies between those of the IMM/PDAF and IMM/MHT. Zaveri *et al* [72] proposed a neural-network algorithm with a IMM filter for multiple maneuvering target tracking, where the assignment weight was evaluated by using the EM algorithm and Hopfield network.

Multiple target tracking: The well-known probabilistic multiple-hypothesis tracker (PMHT) [10, 73, 74, 75] is a preferred multitarget target tracking and association algorithm derived from the application of the EM algorithm. A fundamental difference between the PMHT and other standard tracking approaches is that PMHT assumes that the assignment indices for each measurement are independent random variable. The PMHT algorithm forms an estimate of the unknown model states based on a collection of state observations with uncertain origin, and estimates the model states by maximizing the conditional expectation of the log likelihood with respect to the model to measurement assignments. The computational cost of the PMHT is linear with respect to the number of time steps, measurements and targets.

Multiple detection system target tracking: The multiple detection systems generate multiple measurements simultaneously of the same target via different measurement modes, and the association hypothesis among targets, measurements and measurement modes are unknown. Pulford and Logothetis [76] presented the expectation maximization data association (EMDA) for fixed-interval Kalman smoothing conditioned on the MAP estimation of target-measurement-mode triple association sequence in the EM framework. However, the EMDA is suitable for off-line, batch computation. Lan *et al* [25] proposed the joint multipath data association and state estimation (JMAE) algorithm based on the EM algorithm to obtain the approximate solutions, which carried out the identification of ionospheric mode and measurement association in the E-step, and updated the state estimate in the M-step, where path-conditional state estimation and multipath state fusion are implemented. More recently, Lan *et al* [27] proposed a distributed EM based on consensus filtering (DCEM) to solve the approximate problem posed by JMAE, and they proposed a distributed ECM algorithm (DECM) in [29] to extend the target tracking using OTHR measurements where ionospheric heights are unknown.

3.2.2. Multiple sensors multiple target tracking

Consider the multisensor multitarget tracking in a centralized measurement level fusion, Molnar and Modestino [77] proposed an iterative procedure for time-recursive multitarget/multisensor tracking based on EM algorithm. More specifically, target updates at each time by using an EM based approach that calculates the MAP estimate of the target states, under the assumption of appropriate motion models. The approach uses a Markov random field (MRF) model of the associations between observations and targets and allows for estimation of joint association probabilities without explicit enumeration. Frenkel and Feder [78] investigated the application of EM algorithm to the classical problem of multitarget tracking for a known number of targets, three different schemes, including EM-Newton, EM-Kalman and EM-HMM were proposed based on different optimization criteria. The EM-Newton was a second approximation of the recursive EM algorithm in maximum likelihood criteria, and the EM-Kalman algorithm used EM localization with Kalman tracking in MSE criteria, while the EM-HMM algorithm that used a discrete model for the parameters and a Viterbi search for the optimum parameter sequence in the MAP criteria. These algorithms integrate the localization stage and tracking stage jointly.

The problem of multi-target joint detection and tracking based on EM algorithm by using range and Doppler data from multiple radar platforms was addressed by John *et al* in [79].

Sensor registration and data association are two fundamental problems in multisensor multi-target tracking systems, they actually affect each other. That is, registration requires correctly associated data, and data with sensor biases will result in mistake association. Li and Chen [80] presented a joint data association, registration and fusion algorithm based on the EM framework for multisensor multitarget tracking. More precisely, the target state is regarded as the missing data and estimated in the E-step via KS, and the optimal target-to-measurement association is chosen via the multi-dimensional assignment algorithm, thus the analytical solutions of registration parameters are obtained by maximizing the Q -function. Further, the joint registration and track fusion at the track level was implemented based on the batch and recursive EM algorithms [81]. The EM based extended Kalman smoother (EM-EKS) is developed to perform fusing of the system states and to simultaneously estimate the sensor biases for heterogeneous sensors fusion between radar and electronic support measure (ESM) [82]. A similar approach that results in the EM-IMM-EKF algorithm was found in [83] and it is implemented to solve the joint sensor registration and fusion for cooperative driving in intelligent transportation systems.

3.2.3. Distributed target tracking in sensor networks

The convergence of developments in microelectromechanical system (MEMS), microprocessors and *ad hoc* networking protocols have enabled low-power and low-cost sensor nodes endowed with sensing and processing capabilities to collaborate and achieve large tasks [92]. Distributed estimation and tracking is one of the most fundamental collaborative information processing problems in sensor networks [93], which have many advantages on robustness and scalability particularly in large scale sensor networks over the centralized scheme due to their distributed nature, node redundancy, and the lack of dangers when the fusion center fails.

Distributed processing in sensor networks depends on the level of cooperation allowed between nodes, and strategies may include the incremental strategy, consensus strategy, diffusion strategy and gossip strategy, etc. Nowak proposed a distributed EM algorithm for the joint density estimation and clustering in sensor networks [84], where an incremental approach, which said that a message cycles across every node in the network exactly one time per cycle (*Hamiltonian* cycle). This is a limitation since finding a *Hamiltonian* cycle is an NP-complete problem, and letting the algorithm depend on a single message traveling across the network lacks robustness. Gu [85] proposed a consensus based distributed EM algorithm to handle this difficulty through estimating the global sufficient statistics using local information and neighbors' local information. By using this consensus filter, each node can gradually diffuse its local information over the entire network and asymptotically converge to the global sufficient statistics. However, in a large scale sensor network, the consensus for all the sensor nodes at each iteration can only be achieved with large amount of communications cost. The convergence rate of the consensus algorithm will degrade as the number of sensor nodes increase, if the amount of information exchanged by the nodes does not scale well with their total number. Yang *et al* [86] proposed a diffusion scheme of EM algorithm for Gaussian mixtures in wireless sensor networks (WSNs). At each iteration, the time-varying communication network is modeled as a random graph, and a diffusion-step (D-step) is implemented between E-step and M-step. The diffusion based EM algorithm only needs to update the estimate for each node by single time scale, which can therefore lead to good performance with less communication overhead. The Gossip strategy allows sensors in a network to ignore routing and just exchange data with their nearest neighbors, which

is attractive due to its robustness to unreliable wireless network conditions [94]. The on-line gossip-based distributed EM algorithm for Gaussian mixtures in WSN was proposed in [87].

Different from the centralized joint registration and fusion [80], which carried out at the measurement level for multisensor fusion, the truly distributed joint registration and fusion in sensor network was addressed in [88], where an EM algorithm is developed to perform the track registration, data association and fusion simultaneously at the track level. Moreover, Lan *et al* [28] proposed and compared two centralized EM algorithms and three consensus-based distributed EM algorithms for joint state estimation and identification of sensor networks with unknown inputs. Pereira *et al* [95] proposed a diffusion-based distributed EM algorithm (DB-DEM) for distributed estimation in unreliable sensor networks, where sensors may be subject to data failures and only report noise. The Gaussian mixture modeling for cooperative localization with sensor bias in WSNs was addressed in [96]. By approximating the measurement error distribution with Gaussian mixture model, one centralized and two distributed ECM algorithms were proposed to approximate the MLE of the unknown parameters.

3.2.4. Other joint tracking context

Beside the EM for joint state estimation and identification of unknown parameters, other joint tracking context including joint identification and classification (JIC), joint estimation and localization (JEL), and joint detection and tracking (or track-before-detect, TBD) also benefit from the EM algorithm. The target classification information aids data association, leading to better track purity and track continuity, and thus improves the target performance. Different from the traditional HMM based classification algorithm assuming that a set of training feature sequences are available for each individual HMM, He *et al* [89] proposed an general EM algorithm based joint class identification and target classification algorithm, to deal with the target classification in the case of unknown sequence-to-target-type information. By casting the sensor self-location problem as an parameter estimation problem for hidden Markov Models, Nikolas *et al* [90] proposed an on-line EM algorithm to localize the sensor network simultaneously with target tracking, which is implemented by a distributed Kalman filter and a novel message passing algorithm. The track-before-detect using unthresholded data to perform detection and tracking simultaneously, is desirable for detecting and tracking the low SNR targets. Xia and Liu [91] proposed an EM algorithm for Bayesian TBD with target amplitude fluctuation, in which the fluctuation models are incorporated into the likelihood function, and the average target return amplitude is estimated by the EM algorithm.

Finally, we wish to point out that the EM based algorithm for joint environment parameter identification and target state estimation has also been applied to the applications other than target tracking, for example, the simultaneous localization and mapping (SLAM) [97], cognitive radar [98], and etc. In robotics, the SLAM is the computational problem of constructing or updating a map of an unknown environment while simultaneously keeping track of an agent's location within it, whereas the agent's location is a estimation problem and the environment map is an identification problem. The cognitive radar constitutes a dynamic closed feedback loop encompassing the transmitter, environment and receiver, in order to identify the environmental model parameters and estimate target state simultaneously.

Remark 3.5. *The EM algorithm for joint tracking has many desirable features. It provides a uniform Bayesian framework to handle the situations where system uncertainties are modeled by a hidden Markov process with unknown parameters, such as the uncertainties of unknown input, data association, nonlinear approximation etc. Furthermore, EM based extensions have been*

597 applied to cases where multiple correlated uncertainties are co-exist. In addition, the consensus
 598 based distributed EM algorithm has shown to be effective for distributed estimation in sensor net-
 599 works. Guaranteed convergence and linear computational complexity of the EM techniques are
 600 attractive to many practical target tracking applications. Nevertheless, sometimes, EM method
 601 is difficult to be initialized and estimation outcomes depend on the choice of initial value. It may
 602 converge to a poor locally optimal solution, and the setting of termination conditions is still an
 603 ad hoc in general.

604 4. Examples

605 In this section, we demonstrate the joint tracking applications using the EM method to solve
 606 the OTHR target tracking problem in a multipath environment (*single target tracking*), and the
 607 target collaborative tracking problem in a distributed sensor network with biased measurements
 608 (*multisensor data fusion*), respectively.

609 4.1. Multipath Target Tracking with OTHR Data

610 4.1.1. Problem statement

611 The OTHR exploits a part of the atmosphere that composed of several regions of high ion
 612 density, referred to as layers. Since target echoes may return by a different layer from the one
 613 that transported the transmitted signal, the OTHR is a multipath sensing environment with the
 614 number of possible paths given by the square of the number of layers.

615 The target state in ground coordinates $\mathcal{G}$ at time instant k is defined by $x_k = [R_k, \dot{R}_k, \vartheta_k, \dot{\vartheta}_k]^T$,
 616 which consists of ground range R , range rate $\dot{R}$, bearing ϑ and bearing rate $\dot{\vartheta}$. The discrete-time
 617 state equation is

$$x_k = f(x_{k-1}) + v_k \quad (20)$$

618 where the state transition function f is given; v_k is a zero-mean Gaussian white noise vector with
 619 known covariance Q_k .

620 In OTHR, each received measurement may originate from the underlying target under a partic-
 621 ular propagation mode, or clutters due to false alarm. In other words, any measurement belongs
 622 to the set containing the possible target-originated measurements and clutter-originated false
 623 measurements, denoted $Y_k = \bigcup_{\tau=1}^{\iota} y_k^{\tau} \cup C_k$. Here, C_k is the false measurement where clutters
 624 are independently and uniformly distributed while the clutter number is Poisson distributed, and
 625 each measurement $y_k = [r_k, \dot{r}_k, \zeta_k]^T$ in slant coordinates $\mathcal{R}$ consists of slant range r_k , Doppler $\dot{r}_k$
 626 and azimuth ζ_k . The target-originated measurement y_k via mode τ is randomly available with
 627 known probability p_d^{τ} with its measurement model

$$y_k^{\tau} = h_k^{\tau}(x_k) + w_k^{\tau}, \quad \tau = 1, \dots, \iota \quad (21)$$

628 where h_k^{τ} is the measurement function, w_k^{τ} is a zero-mean Gaussian white noise with known
 629 covariance R_k^{τ} , and ι is the number of possible ionospheric modes. The initial state x_0 is Gaussian
 630 distributed with known parameters. Here, v_k , w_k^{τ} and x_0 are mutually independent.

631 The objective is to obtain the optimal estimation $\hat{x}_k$ of target state x_k under multiple uncer-
 632 tainties of target-to-measurement association a_k and measurement-to-mode association b_k , given
 633 the measurement sets $Y_{1:K}$ up to time K . Under Bayesian inference, this posterior density of x_k
 634 conditioned on $Y_{1:K}$ is written as

$$p(x_k|Y_{1:K}) = \sum_{a_k} \sum_{b_k} p(x_k|a_k, b_k, Y_{1:K})p(a_k, b_k) \quad (22)$$

The Bayesian inference sums over all possible association hypothesis of the hidden variables a_k and b_k . The number of terms in the summation are exponentially increased as the number of possible hypotheses increase. In practical OTHR tracking applications, exact calculation (22) is prohibitively expensive. In our previous work [25], we proposed the JMAE in the EM framework to carry out state estimation and multipath data association identification jointly.

Remark 4.1. The difficulty of tracking a target using OTHR data arises from the uncertain origin of measurements due to multipath propagation, which involves both state estimation x_k and parameter identification $\Theta_k \triangleq \{a_k, b_k\}$. Although the target state x_k is what we are interested in, the parameter Θ_k should be identified accurately. The EM-based iterative optimization is desirable to guarantee the consistence between estimation and identification.

4.1.2. Joint tracking in the EM framework

The JMAE algorithm carries out the identification of ionospheric mode and measurement association in the E-step, where the pseudo-measurement and posterior probabilities of each propagation mode are derived. Meanwhile, the JMAE updates the state estimate in the M-step, where path-conditional state estimates and multipath state fusion are implemented. Due to the iterative optimization, the closed loop between identification and estimation is established, which is desirable in dealing with the coupling of identification and estimation. More importantly, the computational cost of JMAE algorithm is linear given by $O\left(\sum_{k=1}^K (\sum_{\tau=1}^L m_k^\tau)^2\right) + O(Kp^3)$, where m_k^τ is the number of measurement via the τ th mode at time instant k , and p is the dimension of target state. Its information flow is illustrated in Fig.3 and the detailed implementation is given as follows:

Figure 3: The EM framework for joint tracking using OTHR measurements

Step 1: (Modeling) In the EM-based JMAE algorithm, the incomplete-data is OTHR multipath measurements $Y_{1:K}$, missing data is the multipath data association hypothesis $\Theta_{1:K}$, and estimated parameter is the target state $X_{1:K}$.

Step 2: (Joint probability density function) The JMAE algorithm is in MAP sense, and the joint probability density function $p(X_{1:K}, Y_{1:K}, \Theta_{1:K})$ is given by

$$\begin{aligned} \log p(X_{1:K}, Y_{1:K}, \Theta_{1:K}) = & \sum_{k=1}^K \log p(y_k | x_k, a_k, b_k) + \sum_{k=1}^K \log p(x_k | x_{k-1}) + \log p(x_0) \\ & + \sum_{k=1}^K \log p(a_k | b_k) + \sum_{k=1}^K \log p(b_{k-1}^u, b_k^v) + \log p(b_0) \end{aligned} \quad (23)$$

where the measurement-to-mode association b_k obeys a discrete-time s -state Markov chain with known transition probabilities $p(b_{k-1}^u, b_k^v) \triangleq p(b_k = v | b_{k-1} = u)$ and initial probabilities $p(b_0)$.

Step 3: (E-step) Calculate the Q -function $Q(X_{1:K} | X_{1:K}^{(r)})$ and identify the multipath data association $\gamma_k^{(r)}(n, \tau)$ at the r th iteration:

$$\begin{aligned} Q(X_{1:K} | X_{1:K}^{(r)}) = & -\frac{1}{2} \sum_{k=1}^K \sum_{\tau=1}^{\ell} \sum_{n=1}^{m_k^{\tau}} \mathcal{D}(y_k(n) - h^{\tau}(x_k), R_k^{\tau}) \gamma_k^{(r)}(n, \tau) \\ & -\frac{1}{2} \sum_{k=1}^K \mathcal{D}(x_k - f(x_{k-1}, Q_k)) - \frac{1}{2} \mathcal{D}(x_0 - \hat{x}_{0|1:K}, \Sigma_{0,0|1:K}) + C \end{aligned} \quad (24)$$

where the operator $\mathcal{D}(x, P) \triangleq x^T P^{-1} x$, and $\gamma_k^{(r)}(n, \tau) \triangleq p(a_k = n, b_k = \tau | Y_{1:K}, \hat{X}_{1:K}^{(r)})$ is the joint probability density function of a_k and b_k . The quantity C is independent of $X_{1:K}$.

In order to obtain the Q -function, the key is to evaluate the conditional expectation of the joint probability density function $\gamma_k^{(r)}(n, \tau)$, which can be obtained by a HMMs. After that, the pseudo-measurement $\{\tilde{y}_k^{\tau}, \tilde{R}_k^{\tau}\}$ and posterior probabilities $\gamma_k(\tau)$ of each mode τ are given as follows:

$$\tilde{y}_k^{\tau} = \frac{\sum_{n=1}^{m_k^{\tau}} \gamma_k^{(r)}(n, \tau) y_k(n)}{1 - \gamma_k^{(r)}(-1, \tau) - \gamma_k^{(r)}(0, \tau)}, \quad \tilde{R}_k^{\tau} = \frac{R_k^{\tau}}{1 - \gamma_k^{(r)}(-1, \tau) - \gamma_k^{(r)}(0, \tau)} \quad (25)$$

$$\gamma_k^{(r)}(\tau) \triangleq p(a_k = \tau | Y_{1:K}, X_{1:K}^{(r)}) = \sum_{n=-1}^{m_k^{\tau}} \gamma_k^{(r)}(n, \tau) \quad (26)$$

Step 4: (M-step) Maximize the Q -function and estimate the target state $\hat{X}_{1:K}$ as follows:

$$\hat{X}_{k|1:K}^{(r+1)} = \frac{\sum_{\tau=1}^{\ell} \gamma_k^{(r)}(\tau) (\Sigma_{k|1:K}^{\tau})^{-1} \hat{x}_{k|1:K}^{\tau}}{\sum_{\tau=1}^{\ell} \gamma_k^{(r)}(\tau) (\Sigma_{k|1:K}^{\tau})^{-1}}, \quad P_{k|1:K}^{(-1)(r+1)} = \sum_{\tau=1}^{\ell} \gamma_k^{(r)}(\tau) (\Sigma_{k|1:K}^{\tau})^{-1} \quad (27)$$

where $\hat{x}_{k|1:K}^{\tau}$ and $\Sigma_{k|1:K}^{\tau}$ is the state smoothing at time k of mode τ . For a linear or nonlinear dynamic system, this can be carried out by using a fixed interval smoother.

Step 5: (Iteration) The E-M steps are repeated until the values of likelihood functions in two consecutive iterations are close enough or the number of iterations reaches bound.

Step 6: (Outputs) Output the estimation results of path-conditional state estimation $\hat{X}_{1:K}^{\tau}$, multipath track fusion $\hat{X}_{1:K}$, and identification results of data association $\{\tilde{y}_{1:K}^{\tau}, \tilde{R}_{1:K}^{\tau}\}$ and ionosphere mode identification $\gamma_{1:K}(\tau)$.

4.1.3. Simulation results

Consider a single target tracking of OTHR in the clutter environment with an average false measurement density of 400 points per dwell. The number of possible mode $\iota = 4$, and the probability of target detection for each mode τ is $p_d^r = 0.4$, the revisit period is 20s and the target is present over the observation period of 20 dwells. More details about scenario parameters see [25]. We compare the performance of JMAE with that of the well-known MPDA.

The OTHR multipath detections in clutter with four modes are shown in Fig.4. Fig.5 shows the computational cost which linearly increases with respect to the window length. The estimation performance with various number of iterations are shown in Fig.6. As the number of iterations increase, the estimation errors decrease, which shows the convergence of the proposed JMAE. Meanwhile, the JMAE with $r = 0$ is an open-loop smoother, otherwise, it is a closed loop processing. It shows that the closed loop processing is efficient for dealing with the coupling of identification and estimation.

Figure 4: OTHR multipath detection

Figure 5: Computational cost

(a) Estimate error in ground range

(b) Estimate error in bearing

Figure 6: The estimate error with different iteration $r = 0, 1, \dots, 4$

Estimation errors are compared in Fig.7, which shows that the JMAE outperforms MPDA. Such a result is anticipated because the proposed JMAE is an iterative batch processing consists

Figure 7: The performance comparison of JMAE and MPDA

of two smoothers (HMMS in *E-step* and KS in *M-step*), while the MPDA is a recursive filter. Moreover, the MPDA combines multi-path measurements for updating state, which possess an “identification-then-estimation” structure, and its performance decreases when estimation error and identification risk are correlated.

4.2. Target Tracking with Biased Measurements

4.2.1. Problem statement

A fundamental problem in sensor networks concerns how one can fuse the measurements of a target from multiple sensor nodes into a single improved estimate. Before fusion, the sensors must be registered to avoid potential measurement bias. Hence, collaborative tracking in sensor networks involves both the identification of sensor bias and the estimation of target state.

Consider the following discrete-time linear stochastic system for sensor networks.

$$x_k = F_{k-1}x_{k-1} + \Gamma_{k-1}v_{k-1}, \quad (28)$$

$$y_k^j = H_k^j x_k + w_k^j + N_k^j b_k^j, \quad j = 1, \dots, s \quad (29)$$

where $x_k \in \mathbb{R}^n$, $y_k^j \in \mathbb{R}^m$, $b_k^j \in \mathbb{R}^r$ represent the system state, sensor measurement and bias from the j th sensor node with $j = \{1, 2, \dots, s\}$, respectively. Here s is the number of sensor nodes in the network. The matrices F_k , Γ_k , H_k^j , and N_k^j are known with proper dimensions. The process noise $v_k \in \mathbb{R}^p$ and the measurement noise $w_k^j \in \mathbb{R}^m$ are zero-mean white Gaussian noises with known covariances $Q_k > 0$ and $R_k^j > 0$, respectively. The initial state x_0 is Gaussian distributed with known mean $\bar{x}_0$ and associated covariance Σ_0 . Here v_k , w_k^j and x_0 are assumed to be independent, and the sensor bias b_k^j are independent with each other.

The objective is to find the optimal estimate $\hat{x}_k$ of target state x_k and measurement bias estimates $\Theta = \{b_k^j\}_{j=1}^s$ jointly, given measurements from all sensors $Y_{1:K}^{1:s} = \bigcup_{k=1:K}^{j=1:s} y_k^j$. The Bayesian solution is to obtain the posterior density function

$$\begin{aligned}
p(x_k|Y_{1:K}^{1:s}) &= \int_{b_k^1} \cdots \int_{b_k^s} p(x_k, b_k^1, \dots, b_k^s | Y_{1:K}^{1:s}) db_k^1 \cdots db_k^s \\
&= \prod_{j=1}^s \int_{b_k^j} p(x_k | b_k^j, Y_{1:K}^j) p(b_k^j | Y_{1:K}^j) db_k^j
\end{aligned} \tag{30}$$

705 The Bayesian inference requires integrations with respect to the continuous variables
706 $\{b_k^1, \dots, b_k^s\}$, which may not have closed-form analytical solutions, while the dimensionality of
707 the space and the complexity of the integrand may prohibit numerical integration. Thus, the EM
708 based strategies are used to solve the complicated integrations, including centralized methods and
709 distributed methods. In our previous work [28], we compared different EM strategies including
710 centralized and distributed structure to perform the estimation and identification jointly.

711 4.2.2. Joint tracking in the EM framework

712 In this section, two types of centralized EM procedures can be implemented. In addition,
713 we also present three different consensus-based distributed EM approaches. A centralized track
714 fusion based EM (CT-EM) system is illustrated in Fig.8a, where each sensor node carries out
715 EM algorithm to estimate the local state and sensor bias in E-step and M-step, respectively.
716 The global state estimate is obtained at fusion center by fusing all the local state estimates. A
717 centralized measurement fusion based EM (CM-EM) system is shown in Fig.8b, where the fusion
718 center combines measurements from all sensor nodes to perform the joint state estimation and
719 sensor bias identification via the EM algorithm.

720 The distributed expectation maximization (DEM) system is shown in Fig.9, where the joint
721 state estimation and bias identification are performed at local level at each sensor node using
722 standard EM algorithm, and the update of local state estimation is carried out by a consensus
723 filter at global fusion level and sensor bias identification is also updated. Different consensus
724 strategies result to different DEM algorithms.

- 725 • The average consensus based DEM (ADEM) using the local state estimate of the neighbor
726 sensor nodes to update its local state estimate, and hence there is no global state estimate
727 and each sensor node outputs its local state estimate.
- 728 • The iterated consensus based DEM (IDEM) carries out the average consensus filter iterative
729 loop until all the local state estimate reach to a stationary point.
- 730 • The global consensus based DEM (GDEM) consists of two kinds of information. One is
731 the estimated global state information from its neighboring sensor nodes, and the other is
732 the local state estimate.

733 The implementation of the centralized EM and distributed EM are given as follows.

734 **Step 1: (Modeling)** The incomplete-data is the measurements from all sensors $Y_{1:K}^{1:s}$, missing
735 data is the target state $X_{1:K}$, and estimated parameter is the sensor bias $b_{1:K}^{1:s}$.

Step 2: (Likelihood) The EM based methods are in ML sense, and the complete-data log-likelihood for the j th sensor node is given by

$$l^j(X_{1:K}, Y_{1:K}^j | \Theta_{1:K}^j) = \log p(x_0) + \log \sum_{k=1}^K p(x_k | x_{k-1}) + \log \sum_{k=1}^K p(y_k^j | x_k, b_k^j) \tag{31}$$

Figure 8: The framework for joint tracking based on centralized EM

Figure 9: The framework for joint tracking based on distributed EM

Step 3: (E-step) Calculate the Q -function and estimate the target state $X_{1:K}^{i(r)}$ at the r th iteration:

$$\begin{aligned}
 Q(b_{1:K}^j | b_{1:K}^{j(r)}) = & -\frac{1}{2} Tr \left\{ \sum_{k=1}^K Q_k^{-1} \{ C(\hat{x}_{k|1:K} - F_k \hat{x}_{k-1|1:K}) - P_{k,k|1:K} \} \right\} \\
 & -\frac{1}{2} Tr \left\{ \sum_{k=1}^K Q_k^{-1} \{ F_k P_{k,k-1|1:K} + P_{k,k-1|1:K} F_k^T - F_k P_{k-1,k-1|1:K} F_k^T \} \right\} \\
 & -\frac{1}{2} Tr \left\{ \sum_{k=1}^K (R_k^j)^{-1} \{ C(\hat{y}_{k|1:K}^j - H_k \hat{x}_{k|1:K} - N_k \hat{b}_k^j) - H_k P_{k,k|1:K} H_k^T \} \right\}
 \end{aligned} \tag{32}$$

where the operator is defined $C(x) \triangleq x^T x$.

In order to obtain the Q -function, the key is to evaluate the local target state $\hat{X}_{1:K}^{j(r)}$ and its covariance $P_{1:K}^{j(r)}$, which can be obtained by a Kalman smoother. The primary for different EM-based algorithms is how to obtain the fused state estimate $X_{1:K}^{(r)}$ given local state estimate.

- The CM-EM algorithm uses all the measurements and r th sensor bias $\hat{b}_{1:K}^{1:s}(r)$ directly to estimate the target state $\hat{X}_{1:K}^{CM-EM} = \mathbb{E}(X_{1:K}|Y_{1:K}^{1:s}, \hat{b}_{1:K}^{1:s}(r))$.
- The CT-EM algorithm usually uses a weighted fusion algorithm of local estimate results, that is, $\hat{X}_{1:K}^{CT-EM} = \sum_{j=1}^s (P_{1:K}^{j(r)})^{-1} P_{1:K} \hat{X}_{1:K}^{j(r)}$ with $P_{1:K}^{-1} = \sum_{j=1}^s (P_{1:K}^{j(r)})^{-1}$.
- The DEM including ADEM, IDEM and GDEM carries out the consensus filter by using the local state estimate acquired from its neighbor sensor nodes, that is, $\hat{X}_{1:K}^{DEM}(j) = \hat{X}_{1:K}^{DEM}(j) + \eta \left[\hat{X}_{1:K}^{j(r)} - \hat{X}_{1:K}^{DEM}(j) + \sum_{e \in N_j} (\hat{X}_{1:K}^{DEM}(e) - \hat{X}_{1:K}^{DEM}(j)) \right]$, where the η is the weight and N_j represents the neighbors of the j th node.

Step 4: (M-step) Maximize the Q -function and estimate the sensor bias $\hat{b}_{1:K}^{1:s}(r)$ as follows:

$$\hat{b}_k^j = \mathcal{B}_k^j \left(\sum_{k=1}^K (R_k^j)^{-1} (y_k^j - H_k^j \hat{x}_{k|1:K}^{j(r)}) \right), \quad (33)$$

where $\mathcal{B}_k^j = (\mathcal{A}_k^T \mathcal{A}_k^T)^{-1} \mathcal{A}_k^T$, $\mathcal{A}_k = \sum_{k=1}^K (R_k^j)^{-1} N_k^j$, and $\hat{x}_{k|1:K}$ are the fused target state via E-step.

Step 5: (Iteration) The E-M steps are repeated until the values of likelihood functions in two consecutive iterations are close enough or the number of iterations reaches bound.

Step 6: (Outputs) Output the joint estimation results of target state $\hat{X}_{1:K}$ and sensor bias $\hat{b}_{1:K}^{1:s}$.

4.2.3. Simulation results

Consider a target of state $x_k = [\xi_k, \dot{\xi}_k, \eta_k, \dot{\eta}_k]^T$ goes through a sensor network with $s = 10$ agents. The initial position of the target is $(\xi_0, \eta_0) = (0, 0) m$ and the initial velocity is $(\dot{\xi}_0, \dot{\eta}_0) = (15, 0) m/s$. The motion of the target is Constant Velocity (CV). The sampling time T of each agent is $1s$. The whole motion last $60s$. Each agent has an unknown constant bias in the position. System parameters are as follows:

$$F_k = I_2 \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \Gamma_k = I_2 \otimes \begin{bmatrix} T^2/2 \\ T \end{bmatrix}, \text{ and } H_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

the process noise covariance $Q_k = 0.1 I_2 m^2/s^4$. The measurement noise covariance for every agent is $R_k^j = \text{diag}\{1, 1\} m^2$, the biases of each agent is a two dimension random variables $b_k^j \sim U(-100, 100)m$ with uniform distribution, the agents bias control matrix $N_k = I_2$. The initial state covariance $\Sigma_0 = I_2 \otimes \text{diag}(5, 1)$, the iterative terminated $\delta = 10^{-4}$ and the maximum number of iterations $r_{max} = 300$, the window length $l = 5$, the sensor network consists with 10 agents with 10 links, and the weight $W = 0.02$.

Fig.10 shows the target trajectory and the graph of sensor networks. Fig.11 shows the estimation performance of the target state error and the average measurements biases error over all the sensors. All the EM based strategies can output the target state estimation and sensor measurement biases jointly, and the performance of the state estimation and biases identification are affect with each other. As the time sequence increase, all the estimation errors reach to a stationary point. Among all the EM based strategies, the DEM algorithms have better performance than the CT-EM algorithm, the IDEM has better performance than the GDEM and ADEM, and the

Figure 10: The scenario of network graph and target trajectory

Figure 11: The estimation performance of state and bias

766 ADEM is the worst. This is mainly because the ADEM is just average over the whole network,
 767 and its performance decreases when the sparse networks. The IDEM algorithm guarantees all
 768 the local state estimate consensus to the stationary point at every iteration in the EM framework
 769 by using the iterative strategy, while the GDEM takes a long time to reach consistency.

770 5. Conclusion and discussion

771 In this paper, an overview of the approaches in literature for joint tracking using EM algorithm
 772 is presented. We show that a general multi-sensor multi-target tracking system involves both
 773 target state estimation and system identification problems. The interdependence between estima-
 774 tion and identification add additional difficulty for an efficient solution to this tracking problem.
 775 The joint tracking framework, which simultaneously takes both estimation error and identifica-
 776 tion risk of target state into account, can effectively solve this kind of target tracking problem
 777 using the EM algorithm. On the other hand, this target tracking problem can be treated as the

state estimation using *incomplete data*, that is, some unknown parameters need to be identified before we can estimate the underlying target state. The EM algorithm is shown being an effective method to solve a range of joint estimation and identification problems as we demonstrated by the joint estimation and identification examples in the paper. Nevertheless, many important issues on joint estimation and identification using EM algorithm are continuously investigated.

- *Control of EM convergence rate:* Dempster et al. show that the rate of convergence of the EM algorithm is linear and the rate depends on proportion of information in the observed data. Thus in comparison to the formulated complete-data problem, if a large portion of data is missing, convergence can be quite slow. Unfortunately, the target tracking is suffering from the challenges of dense target environment, rapidly maneuvering targets, or complex signal propagation environments. The variational Bayesian (VB) [99, 100, 101, 102, 103] methodology is an iterative approach that is gaining popularity to ameliorate this shortcoming of the EM algorithm. This methodology allows inference in the case of complex graphical models which also guarantees local convergence, that in certain cases provide significant improvements as compared to simpler ones that can be solved via the EM. More details see the survey of VB algorithm [104, 105, 31]. Most recently, the VB for target tracking has been addressed [106]. Perhaps Simo *et al* [107] firstly used the VB algorithm to deal with the target tracking under the unknown measurement noise covariance, and developed the variational Bayesian adaptive Metropolis (VBAM) algorithm [108, 109] to the nonlinear case. Subsequently, the IMM-VB [110, 111], PHD-VB [112, 113] algorithms are proposed for the extension of unknown measurement noise covariance to the jump Markov system and multiple target tracking system, respectively. The VB algorithm for data association problem was presented in [114, 115]. Tinne *et al* [116] proposed the VB algorithm for joint clustering and tracking in multitarget tracking with Merged measurements. Williams [117] proposed an VB based multi-Bernoulli filter for multitarget tracking with the missing data being the correspondence of Bernoulli components.
- *Convexity of the objective function for EM:* In some joint estimation and identification applications, such as joint target tracking and classification [118, 119], or joint detection and tracking [120, 121], the EM performance is also assessed by the Bayesian risk, which consists of both estimation risk and identification risk [122, 123]. In general, the risk function is non-convex, and how to use the EM algorithm in the non-convex remains an open problem.
- *EM for practical applications:* The computational complexity and accuracy are two major measures in the practical algorithm implementation. In the EM framework, there are some *ad hoc* parameters including the sliding/batch window length, iteration termination threshold, etc. It is necessary to determine an on-line performance evaluation strategy for optimizing those parameters, and obtain a trade-off between computational burden and processing accuracy. Meanwhile, construct appropriate constraints in the practical applications is useful, since the intuitive posterior constraints can greatly improve the performance of standard EM algorithm and be competitive with more complex, intractable models.

6. Acknowledgment

This work was supported by National Science Foundation of China under Grant 61501378, 61135001, 61374023, 61374159, 61503305, and the Fundamental Research Funds for the Central Universities under Grant G2015KY0801.

7. Reference

- [1] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with Application to Tracking and Navigation: Theory Algorithms and Software*. New York: Wiley, 2001.
- [2] D. Hall and L. James. *Handbook of Multisensor Data Fusion*. CRC Press, 2001.
- [3] M. Persson and G. Rigas. Complexity: the dark side of network-centric warfare. *Cognition Technology and Work*, 16(1):103 – 115, 2014.
- [4] X.R. Li and V.P. Jilkov. Survey of maneuvering target tracking. Part V: Multiple-model methods. *IEEE Transactions on Aerospace and Electronic Systems*, 41(4):1255 – 1321, 2005.
- [5] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401 – 422, 2004.
- [6] 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*, 50(2):174 – 188, 2002.
- [7] G. Terejanu, P. Singla, T. Singh, and P.D. Scott. Adaptive Gaussian sum filter for nonlinear Bayesian estimation. *IEEE Transactions on Automatic Control*, 56(9):2151–2156, 2011.
- [8] S.S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19(1):5 – 18, 2004.
- [9] D.B. Reid. Algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6):843 – 854, 1979.
- [10] P. Willett, Y.H. Ruan, and R. Streit. PMHT: Problems and some solutions. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):738 – 754, 2002.
- [11] T.E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8(3):173 – 184, 1983.
- [12] Q. Yu and G. Medioni. Multiple-target tracking by spatiotemporal monte carlo Markov chain data association. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2196 – 2210, 2009.
- [13] S. Särkkä, A. Vehtari, and J. Lampinen. Rao-blackwellized particle filter for multiple target tracking. *Information Fusion*, 8(1):2 – 15, 2007.
- [14] R. Mahler. “Statistics 102” for multisource-multitarget detection and tracking. *IEEE Journal on Selected Topics in Signal Processing*, 7(3):376 – 389, 2013.
- [15] X.R. Li and V.P. Jilkov. A survey of maneuvering target tracking: Approximation techniques for nonlinear filtering. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 5428, pages 537 – 550, 2004.
- [16] G.W. Pulford. Taxonomy of multiple target tracking methods. *IEE Proceedings: Radar, Sonar and Navigation*, 152(5):291 – 304, 2005.
- [17] I.J. Cox. Review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10(1):53 – 66, 1993.
- [18] D. Smith and S. Singh. Approaches to multisensor data fusion in target tracking: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 18(12):1696 – 1710, 2006.
- [19] B. Khaleghi, A. Khamis, F.O. Karray, and S.N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28 – 44, 2013.
- [20] S. Lauro, G. Jesús, and L. James. Context-based information fusion: A survey and discussion. *Information Fusion*, 25:16 – 31, 2015.
- [21] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1 – 38, 1977.
- [22] X.R. Li. Optimal Bayes joint decision and estimation. In *10th International Conference on Information Fusion*, pages 1 – 8, July 2007.
- [23] W. Cao, J. Lan, and X.R. Li. Conditional joint decision and estimation with application to joint tracking and classification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–13, 2015.
- [24] X.R. Li, M. Yang, and J.F. Ru. Joint tracking and classification based on Bayes joint decision and estimation. In *10th International Conference on Information Fusion*, pages 1–8, July 2007.
- [25] H. Lan, Y. Liang, Q. Pan, F. Yang, and C. Guan. An EM algorithm for multipath state estimation in OTHR target tracking. *IEEE Transactions on Signal Processing*, 62(11):2814 – 2826, 2014.
- [26] H. Lan, Y. Liang, F. Yang, Z.F. Wang, and Q. Pan. Joint estimation and identification for stochastic systems with unknown inputs. *IET Control Theory and Applications*, 7(10):1377 – 1386, 2013.
- [27] H. Lan, Y. Liang, Z.F. Wang, F. Yang, and Q. Pan. A distributed expectation-maximization algorithm for OTHR multipath target tracking. In *17th International Conference on Information Fusion*, pages 1 – 8, 2014.
- [28] H. Lan, A.N. Bishop, and Q. Pan. Distributed joint estimation and identification for sensor networks with unknown inputs. In *2014 IEEE 9th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Conference Proceedings*, pages 1 – 6, 2014.

- [29] H. Lan, Z.F. Wang, F. Yang, and Q. Pan. Joint OTHR multipath state estimation with unknown ionospheric heights. In *16th International Radar Symposium (IRS)*, pages 753 – 758, 2015.
- [30] D. Tzikas, A. Likas, and N. Galatsanos. The variational approximation for Bayesian inference: Lift after the EM algorithm. *IEEE Signal Processing Magazine*, 25(6):131 – 146, 2008.
- [31] Bishop C.M. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [32] E. Cosme, J. Verron, P. Brasseur, J. Blum, and D. Auroux. Smoothing problems in a Bayesian framework and their linear Gaussian solutions. *Monthly Weather Review*, 140(2):683 – 695, 2012.
- [33] M.R. Gupta and Y.H. Chen. Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, 4(3):223 – 296, 2010.
- [34] M.J. Geoffrey and K. Thriyambakam. *The EM and Its Extension*. John Wiley & Sons., 2008.
- [35] C. F. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95 – 103, 1983.
- [36] S. Balakrishnan, M.J. Wainwright, and B. Yu. Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Eprint Arxiv*, 2014.
- [37] S. Chretien and A.O. Hero. On EM algorithms and their proximal generalizations. *ESAIM - Probability and Statistics*, 12:308 – 326, 2008.
- [38] P. Tseng. An analysis of the EM algorithm and entropy-like proximal point methods. *Mathematics of Operations Research*, 29(1):27 – 44, 2004.
- [39] C. Bouveyron and C. Brunet. Theoretical and practical considerations on the convergence properties of the Fisher-EM algorithm. *Journal of Multivariate Analysis*, 109:29 – 41, 2012.
- [40] J. Graca, K. Ganchev, and B. Taskar. Expectation maximization and posterior constraints. In *Neural Information Processing Systems Conference 20*, pages 1 – 8, 2007.
- [41] J. Ahn and J. Oh. A constrained EM algorithm for principal component analysis. *Neural Computation*, 15(1):57 – 65, 2003.
- [42] X.L. Meng and D.B. Rubin. On the global and component wise rates of convergence of the EM algorithm. *Linear Algebra and its Applications*, 19(9):413 – 425, 1994.
- [43] V. Melnykov and I. Melnykov. Initializing the EM algorithm in Gaussian mixture models with an unknown number of components. *Computational Statistics and Data Analysis*, 56(6):1381 – 1395, 2012.
- [44] C. Fraley. Algorithms for model-based Gaussian hierarchical clustering. *SIAM Journal of Scientific Computing*, 20(1):270 – 281, 1998.
- [45] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis*, 41(3-4):561 – 575, 2003.
- [46] R. Maitra. Initializing partition-optimization algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6(1):144 – 157, 2009.
- [47] A. Roche. EM algorithm and variants: an informal tutorial. *ArXiv e-prints*, 2011.
- [48] J.A. Fessler and A.O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, 42(10):2664 – 2677, 1994.
- [49] X.L. Meng and D.B. Rubin. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2):267 – 278, 1993.
- [50] Liu C.H. and D.B. Rubin. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81(4):633 – 648, 1994.
- [51] Y.X. He and C.H. Liu. The dynamic expectation-conditional maximization either algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):313 – 336, 2012.
- [52] C.G. Greg and A.T. Martin. A monte carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699 – 704, 1990.
- [53] C.H. Liu, D.B. Rubin, and Y.N. Wu. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, 85(4):755 – 770, 1998.
- [54] J. Mortaza and I. J. Robert. Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association*, 88(421):221 – 228, 1993.
- [55] R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, volume 89, pages 355 – 368. Springer Netherlands, 1998.
- [56] W. Kowalczyk and N. Vlassis. Newscast EM. In *Advances in Neural Information Processing Systems 17*, pages 713 – 720. MIT Press, 2005.
- [57] O. Cappé and E. Moulines. On-line expectation-maximization algorithm for latent data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):593 – 613, 2009.
- [58] T. Denœux. Maximum likelihood from fuzzy data using the EM algorithm. *Fuzzy Sets and Systems*, 183:72 – 91, 2011.
- [59] T. Denœux. Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):119 – 130, 2013.

- [60] J.M Jiang, N. Thuan, and J. S. Rao. The E-MS algorithm: Model selection with incomplete data. *Journal of the American Statistical Association*, 2014.
- [61] R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253 – 264, 1982.
- [62] Z. Ghahramani and E. H. Geoffrey. Parameter estimation for linear dynamical systems. Technical report, University of Toronto, Department of Computer Science, 1996.
- [63] A. Zia, T. Kirubarajan, J.P. Reilly, D. Yee, K. Punithakumar, and S. Shirani. An EM algorithm for nonlinear state estimation with model uncertainties. *IEEE Transactions on Signal Processing*, 56(3):921 – 936, 2008.
- [64] H. Karimi and K.B. McAuley. An approximate expectation maximization algorithm for estimating parameters, noise variances, and stochastic disturbance intensities in nonlinear dynamic models. *Industrial and Engineering Chemistry Research*, 52(51):18303 – 18323, 2013.
- [65] M. Lei, C.Z. Han, and P.Z. Liu. Expectation maximization (EM) algorithm-based nonlinear target tracking with adaptive state transition matrix and noise covariance. In *FUSION 2007 - 2007 10th International Conference on Information Fusion*, pages 212 – 218, 2007.
- [66] E. Özkan, F. Lindsten, C. Fritsche, and F. Gustafsson. Recursive maximum likelihood identification of jump Markov nonlinear systems. *IEEE Transactions on Signal Processing*, 63(3):754 – 765, 2015.
- [67] A. Logothetis and V. Krishnamurthy. Expectation maximization algorithms for MAP estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 47(8):2139–2156, 1999.
- [68] L.A. Johnston and V. Krishnamurthy. An improvement to the interacting multiple model (IMM) algorithm. *IEEE Transactions on Signal Processing*, 49(12):2909 – 2923, 2001.
- [69] G.W. Pulford and Barbara L.F. MAP estimation of target manoeuvre sequence with the expectation-maximization algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 38(2):367 – 377, 2002.
- [70] A. Logothetis, V. Krishnamurthy, and J. Holst. A Bayesian EM algorithm for optimal tracking of a maneuvering target in clutter. *Signal Processing*, 82(3):473 – 490, 2002.
- [71] Y.H. Ruan and P. Willett. Multiple model PMHT and its application to the benchmark radar tracking problem. *IEEE Transactions on Aerospace and Electronic Systems*, 40(4):1337 – 1350, 2004.
- [72] M.A. Zaveri, S.N. Merchant, and U.B. Desai. Robust neural-network-based data association and multiple model-based tracking of multiple point targets. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3):337 – 351, 2007.
- [73] R.L. Streit. *Poisson Point Processes: Imaging, Tracking, and Sensing*. Springer, 2010.
- [74] T. Long, L. Zheng, X.L. Chen, Y. Li, and T. Zeng. Improved probabilistic multi-hypothesis tracker for multiple target tracking with switching attribute states. *IEEE Transactions on Signal Processing*, 59(12):5721 – 5733, 2011.
- [75] M. Wieneke and W. Koch. A PMHT approach for extended objects and object groups. *IEEE Transactions on Aerospace and Electronic Systems*, 48(3):2349 – 2370, 2012.
- [76] G.W. Pulford and A. Logothetis. An expectation-maximization tracker for multiple observations of a single target in clutter. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 4997–5003, 1997.
- [77] K.J. Molnar and J.W. Modestino. Application of the EM algorithm for the multitarget/multisensor tracking problem. *IEEE Transactions on Signal Processing*, 46(1):115 – 129, 1998.
- [78] L. Frenkel and M. Feder. Recursive expectation-maximization (EM) algorithms for time-varying parameters with applications to multiple target tracking. *IEEE Transactions on Signal Processing*, 47(2):306 – 320, 1999.
- [79] R. Deming, J. Schindler, and L. Perlovsky. Multi-target/multi-sensor tracking using only range and doppler measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 45(2):593 – 611, 2009.
- [80] Z.H. Li, S.Y. Chen, H. Leung, and E. Bosse. Joint data association, registration, and fusion using EM-KF. *IEEE Transactions on Aerospace and Electronic Systems*, 46(2):496 – 507, 2010.
- [81] D.L. Huang, H. Leung, and E. Bosse. A pseudo-measurement approach to simultaneous registration and track fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 48(3):2315 – 2331, 2012.
- [82] Z.H. Li, H. Leung, and L. Tao. Simultaneous registration and fusion of radar and ESM by EM-EKS. In *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, pages 1130 – 1134, Jinan, China, 2010.
- [83] D.L. Huang and H. Leung. An expectation-maximization-based interacting multiple model approach for cooperative driving systems. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):206 – 228, 2005.
- [84] R.D. Nowak. Distributed EM algorithms for density estimation and clustering in sensor networks. *IEEE Transactions on Signal Processing*, 51(8):2245 – 2253, 2003.
- [85] D.B. Gu. Distributed EM algorithm for Gaussian mixtures in sensor networks. *IEEE Transactions on Neural Networks*, 19(7):1154 – 1166, 2008.
- [86] Y. Weng, W.D. Xiao, and L.H. Xie. Diffusion-based EM algorithm for distributed estimation of Gaussian mixtures in wireless sensor networks. *Sensors*, 11(6):6297 – 6316, 2011.
- [87] G. Morral, P. Bianchi, and J. Jakubowicz. On-line gossip-based distributed expectation maximization algorithm.

- In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 305 – 308, 2012.
- [88] H. Zhu, L. Henry, and K.V. Yuen. A joint data association, registration, and fusion approach for distributed tracking. *Information Sciences*, 324:186 – 196, 2015.
- [89] X.F. He, R. Tharmarasa, T. Kirubarajan, A.L. Jousselme, and P. Valin. Joint class identification and target classification using multiple HMMs. *IEEE Transactions on Aerospace and Electronic Systems*, 50(2):1269 – 1282, 2014.
- [90] N.K. Kantas, S.S. Singh, and A. Doucet. Distributed maximum likelihood for simultaneous self-localization and tracking in sensor networks. *IEEE Transactions on Signal Processing*, 60(10):5038 – 5047, 2012.
- [91] S. Z. Xia and H. W. Liu. Bayesian track-before-detect algorithm with target amplitude fluctuation based on expectation-maximisation estimation. *IET Radar Sonar and Navigation*, 6(8):719 – 728, 2012.
- [92] J. Liu, M. Chu, and J.E. Reich. Multitarget tracking in distributed sensor networks. *IEEE Signal Processing Magazine*, 24(3):36 – 46, 2007.
- [93] I.F. Akyildiz, W.L. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102 – 105, 2002.
- [94] A.G. Dimakis, S. Kar, J.M.F. Moura, M.G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847 – 1864, 2010.
- [95] S.S. Pereira, L.V. Roberto, and P.Z. Alba. A diffusion-based EM algorithm for distributed estimation in unreliable sensor networks. *IEEE Signal Processing Letters*, 20(6):595 – 598, 2013.
- [96] Y. Feng, C. Fritsche, J. Di, F. Gustafsson, and A.M. Zoubir. Cooperative localization in WSNs using Gaussian mixture modeling: Distributed ECM algorithms. *IEEE Transactions on Signal Processing*, 63(6):1448 – 1463, 2015.
- [97] D.W. Hugh and T. Bailey. Simultaneous localization and mapping: Part I. *IEEE Robotics and Automation Magazine*, 13(2):99 – 108, 2006.
- [98] S. Haykin. Cognitive radar: A way of the future. *IEEE Signal Processing Magazine*, 23(1):30 – 40, 2006.
- [99] M. Beal. *The Variational algorithms for approximation Bayesian inference*. PhD thesis, University of Cambridge, 1998.
- [100] J. Winn. *Variational message passing and its applications*. PhD thesis, University of Cambridge, 2003.
- [101] J. Winn and C.M. Bishop. Variational message passing. *Journal of Machine Learning Research*, 6:661 – 694, 2005.
- [102] M.D. Hoffman, D.M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303 – 1347, 2013.
- [103] J. Sung, Z. Ghahramani, and S.Y. Bang. Latent-space variational bayes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2236 – 2242, 2008.
- [104] C.W. Fox and S.J. Roberts. A tutorial on variational Bayesian inference. *Artificial Intelligence Review*, 38(2):85 – 95, 2012.
- [105] S.L. Sun. A review of deterministic approximate inference techniques for Bayesian machine learning. *Neural Computing and Applications*, 23(7-8):2039 – 2050, 2013.
- [106] V. midl and A. Quinn. Variational Bayesian filtering. *IEEE Transactions on Signal Processing*, 56(10):5020 – 5030, 2008.
- [107] S. Särkkä and A. Nummenmaa. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control*, 54(3):596 – 600, 2009.
- [108] A. Juha, S. Särkkä, and P. Robert. Gaussian filtering and variational approximations for Bayesian smoothing in continuous-discrete stochastic dynamic systems. *Signal Processing*, 111:124 – 136, 2015.
- [109] I. S. Mbalawata, S. Särkkä, M. Vihola, and H. Haario. Adaptive metropolis algorithm using variational Bayesian adaptive Kalman filter. *Computational Statistics and Data Analysis*, 83:101 – 115, 2015.
- [110] W. Li and Y. Jia. State estimation for jump Markov linear systems by variational Bayesian approximation. *IET Control Theory and Applications*, 6(2):319 – 326, 2012.
- [111] C. Shen, D.J. Xu, W. Huang, and F. Shen. An interacting multiple model approach for state estimation with non-Gaussian noise using a variational Bayesian method. *Asian Journal of Control*, 17(4):1424 – 1434, 2015.
- [112] W.L. Li, Y.M. Jia, J.P. Du, and J. Zhang. PHD filter for multi-target tracking by variational Bayesian approximation. In *Proceedings of the IEEE Conference on Decision and Control*, pages 7815 – 7820, 2013.
- [113] J.L. Yang and H.W. Ge. An improved multi-target tracking algorithm based on CBMeMber filter and variational Bayesian approximation. *Signal Processing*, 93(9):2510 – 2515, 2013.
- [114] L.G. Miguel, V. V. Steven, and N.D. Lawrence. Overlapping mixtures of Gaussian processes for the data association problem. *Pattern Recognition*, 45(4):1386 – 1395, 2012.
- [115] D. T. Ryan, B. Steven, and A. Bhargav. A complete variational tracker. In *Advances in Neural Information Processing Systems 27*, pages 496 – 504. Curran Associates, Inc., 2014.
- [116] D. Tinne., B. Herman, and D.S. Joris. Shape-based online multitarget tracking and detection for targets causing multiple measurements: Variational Bayesian clustering and lossless data association. *IEEE Transactions on*

- 1056 *Pattern Analysis and Machine Intelligence*, 33(12):2477 – 2491, 2011.
- 1057 [117] J.L. Williams. An efficient, variational approximation of the best fitting multi-Bernoulli filter. *IEEE Transactions*
1058 *on Signal Processing*, 63(1):258 – 273, 2015.
- 1059 [118] T. Vercauteren, D. Guo, and X.D. Wang. Joint multiple target tracking and classification in collaborative sensor
1060 networks. *IEEE Journal on Selected Areas in Communications*, 23(4):714 – 723, 2004.
- 1061 [119] S. Challa and G.W. Pulford. Joint target tracking and classification using radar and ESM sensors. *IEEE Transac-*
1062 *tions on Aerospace and Electronic Systems*, 37(3):1039 – 1055, 2001.
- 1063 [120] B.T. Vo, D. Clark, B.N. Vo, and B. Ristic. Bernoulli forward-backward smoothing for joint target detection and
1064 tracking. *IEEE Transactions on Signal Processing*, 59(9):4473 – 4477, 2011.
- 1065 [121] B.T. Vo, C. M. See, N. Ma, and W.T. Ng. Multi-sensor joint detection and tracking with the Bernoulli filter. *IEEE*
1066 *Transactions on Aerospace and Electronic Systems*, 48(2):1385 – 1402, 2012.
- 1067 [122] Y. Liu. *Estimation, Decision and Applications to Target Tracking*. PhD thesis, University of New Orleans, 2013.
- 1068 [123] M. Yang. *When Decision Meets Estimation: Theory and Applications*. PhD thesis, University of New Orleans,
1069 2007.