LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-799417

Probabilistic Data Association for Orbital-Element Estimation Using Expectation-Maximization Algorithms

J. Bernstein

December 13, 2019

Journal of Aerospace Information Systems

Disclaimer

Probabilistic Data Association for Orbital-Element Estimation Using Multi-Stage Expectation-Maximization

Jason Bernstein *
Lawrence Livermore National Laboratory, Livermore, CA, 94550

LLNL-JRNL-799417

Tracking space objects is important for managing space traffic and predicting collisions, but is difficult in part due to data association and orbit model uncertainty. Expectation-Maximization (EM) is a commonly used tracking method that has not been widely considered for tracking space objects. The technique consists of iteratively computing data association probabilities with a set of current element estimates, and updating estimates of the elements by solving a non-linear weighted least squares regression problem where the weights are the data association probabilities. This paper demonstrates the use of EM for probabilistic data association and orbital-element estimation by applying the technique to simulated data from two angles-only tracking scenarios. In both scenarios, EM provides correct data associations and accurate maximum likelihood estimates of orbital elements. One scenario considers tracking a single object in clutter and quantifies the improvement of the orbital-element estimates and data associations as the detection probability increases. However, standard application of EM requires knowing the number of objects or may fail when a large number of objects are present. To address these issues, this paper employs a multi-stage version of EM that is applicable when there are a large and possibly unknown number of objects.

Nomenclature

J	=	Number of objects in mixture model
λ	=	Concentration parameter (rad^{-2})
π_j	=	Prior data association probability for jth object
Θ_j	=	Orbital elements for jth object
ξ	=	Mixture model parameters to be estimated
n	=	Number of measurements
t	=	Time (s)

*Statistician, Computational Engineering Division, bernstein8@llnl.gov.

t_i	=	ith measurement time (s)
$\phi_i = (\phi_{i,1}, \phi_{i,2})$	=	ith angle measurement (rad)
Y_i	=	Unit-vector form of ith angle measurement
y_i	=	Corresponding index variable for unit vectors
X_i	=	ith latent data association variable
$p_\xi(y_i)$	=	Distribution of a random variable Y_i
$p_\xi(x_i)$	=	Distribution of a random variable X_i
$p_\xi(y_i \vert X_i = j)$	=	Conditional distribution of ith measurement from jth orbit
$r_O(t)$	=	Position of observer
$r(t;\Theta_j)$	=	Position of jth object
$w_{ij}^{(h)}$	=	Data association probability
Subscripts		
i	=	Measurement index
j	=	Orbit or clutter model index
Superscripts		
h	=	EM algorithm iteration index

I. Introduction

Space object tracking is required to manage space traffic [1], predict collisions [2], and attain Space Situational Awareness (SSA). However, this task is complicated by non-object measurements referred to as clutter, data association and orbit model uncertainty, sparse data, and orbital motion being non-linear [3]. Recall that data association uncertainty in the tracking context refers to not knowing which measurements belong to which objects or are clutter. Several techniques are currently used to track space objects, and are discussed later in this section, but new techniques may be required in the future. Expectation-Maximization (EM) appears to have only recently been used for space object tracking [4–9], but its generality and flexibility suggest it may be adapted to meet future space tracking needs.

A particular need is the ability to track a large number of space objects simultaneously. Previous work that considers tracking space objects by fitting mixture models with EM demonstrate the approach on three or fewer objects [6, 8, 9]. However, the standard EM approach to mixture model estimation may not scale-well with the number of objects due to the high-dimensionality of the parameter space. This paper considers an extension of the general EM approach for mixture model estimation, referred to as multi-stage EM, that is shown in a simulation to enable data association and orbit determination for more objects than standard EM.

The multiple space object tracking problem considered here is summarized for context. Data are assumed to have

been collected from clutter or one or more objects at different times, but their associations are not known. The tracking problem is to associate the measurements to the different objects or clutter and to determine, or estimate, the orbits of the objects; these tasks are referred to as data association and orbit determination, respectively, though it is also common to refer to data association as track linkage or uncorrelated track (UCT) linkage. Data association and orbit determination are coupled in that orbit determination is simplified given the data associations, and similarly data association is simplified given the orbits. Variations of the orbit, detection, or measurement model can be made, but these do not fundamentally change the statistical tracking problem.

Literature reviews on multiple space object tracking can be found in [3, 10], and more general reviews on multiple object tracking include [11–17]. Current tracking or space object tracking approaches include joint probabilistic data association (JPDA) [18, 19], maximum likelihood PDA [20], multiple hypothesis tracking (MHT) [21–24], random finite set (RFS) methods [25–29], multitarget intensity filters [30], the Distinguishable and Independent Stochastic Populations (DISP) filter [31] and the related method [32], and reversible jump Markov Chain Monte Carlo (MCMC) [33]. The methods are typically filter-based and designed to account for data association ambiguity and an unknown number of objects, and are distinguished by the underlying statistical models they assume or their implementation. While the approaches demonstrate good performance in challenging tracking scenarios, many are not as well-known or known as broadly as Expectation-Maximization (EM). Hence, a goal of this paper is to increase interest in EM as a space object tracking technique and to interest a broader range of researchers in this problem as well.

EM is often used for multiple object tracking and is applied to a variety of models and tracking scenarios in [15, 34–45]. Much of the work on tracking with EM considers the probabilistic multiple hypothesis tracking (PMHT) technique [46], which uses a filter for state prediction and has several variations in the literature. [47] and [48] review the literature on EM and PMHT for tracking, respectively. In general, EM can be used to compute maximum likelihood estimates of motion model parameters or object states and data association probabilities under those estimates. The approach tends to be applicable when there is a notion of missing data, and so is well-suited for tracking problems where the data associations are not known. For estimating parameters in the mixture model formulation of the multiple object tracking problem, EM is sometimes called maximum likelihood PMHT, or ML-PMHT [49]. EM is also used in particle filter [50] or RFS [29, 51] methods to estimate Gaussian mixture models related to the state of an object. In contrast, this paper represents the states of objects deterministically and uses EM to simultaneously estimate the elements for each of the objects.

Despite EM being a well-established approach to solving missing data problems, the method and its variant PMHT do not appear to be commonly used for tracking space objects. PMHT is applied in [4] and [7] to decametric radar and event based sensor data, respectively, and to track-before-detect in [5]. In [48, Sec. 12.5], PMHT is applied with random matrices to track a large number of stars and satellites. [8] use EM for data association and tracking a maneuvering object assuming a mixture model of spherical distributions, but do not discuss estimating orbital elements with EM. [6]

and [9] apply different EM algorithms for data association and orbit determination to Gaussian, or similar to Gaussian, mixture models of angles-only data; their approaches differ partly due to the former linearizing the motion model to obtain a least-squares update of the state of the objects. In general, there is some variation in how EM has been implemented for mixture models in this domain, so a goal of this paper is to provide a general presentation that refines the technique.

In particular, the mixture model, EM algorithms, and results presented here differ from those previously considered for space object tracking in several ways. For example, angle measurements are modeled as realizations of the spherical normal distribution [52] in order to avoid potential problems with modeling angle measurements as Gaussian; the Gaussian error model may not describe angle data well since angles describe points on the unit sphere. A clutter model is also included as a possible measurement model, which can help facilitate linkage as discussed in Section II. Additionally, the approach considered here does not rely on linearizing the orbit model and so avoids errors from this approximation. These points are addressed in different parts of the literature, but they are considered together here for completeness.

The main contribution of this paper is a consideration of a multi-stage EM procedure for orbit determination in the mixture model formulation of the multiple space object tracking problem. The procedure is an iterative application of EM, in which measurements unambiguously linked to an orbit model in one implementation or stage of EM are set-aside and EM is reapplied to the remaining measurements. Multi-stage EM is shown to determine a greater number of orbits than the standard EM approach, referred to here as EM or single-stage EM depending on the context, when there are tens of objects. Alternative multi-stage tracking procedures include multistage maximization [13, Sec. 10.6] and multi-stage MHT [53], and [49] consider a similar iterative application of EM to linear, Gaussian mixture models.

Convergence and estimation properties of orbital-element estimates obtained with single-stage EM are also illustrated in two distinct tracking scenarios. Specifically, numerical simulations demonstrate that EM produces maximum likelihood estimates of orbital elements and show the decrease in estimator bias as the detection probability increases. The Rand index statistic [54] from cluster analysis is also employed to show that the approach correctly associates measurements with objects and identifies clutter. The presentation is intended to be general enough that the EM algorithms presented here can be modified to apply to other tracking scenarios, orbit models, or data types.

The paper is organized as follows. A general description of the space object tracking problem is given in Section II and a simulated data set is provided as a motivating example. Section III presents a mixture model formulation of the multiple object tracking problem that characterizes the data association model, object state, and measurement uncertainties. Section IV describes the EM algorithm for fitting this mixture model and describes the multi-stage EM extension that is applicable when the number of objects is large or unknown. EM is shown to provide correct data associations and accurate maximum likelihood estimates of orbital elements for the presented data in Section V.A; Section V.B illustrates multi-stage EM on similarly simulated data but with more objects. Section VI presents results for

a different tracking scenario in which a single object is tracked in clutter, but here the focus is on the performance of EM over multiple data sets; the mixture model and EM algorithm for this scenario are given in the appendix. Section VII considers limitations of EM, the computational cost of EM, and briefly discusses EM for the more general state space model formulation of the multiple object tracking problem, which allows for tracking objects while accounting for process noise. The paper is summarized and future research directions are suggested in Section VIII.

II. Problem Description

A qualitative description of the tracking problem considered here is now given before describing the formal mixture model framework and EM approach for its solution. A single observer is assumed to collect angle measurements over time of one or more objects. The association of the measurements to the objects is not known, but some prior information, such as the range of the elements, may be available. Measurements can also be clutter, or false alarms, that do not correspond to any of the objects. Only one measurement is permitted at each time for now; Section VI considers an example where multiple measurements can be available at each time. The objects are assumed to follow deterministic orbits parameterized by a set of elements whose values are not known.

The goals are to link the measurements to the objects and to estimate the elements for each orbit. More specifically, probabilistic data associations are desired that describe the uncertainty in associations between objects and measurements, and maximum likelihood estimates of orbital elements are sought that take data association uncertainty into account. This paper focuses on the use of EM for data association and point estimation of orbital elements, but discusses approximations for the distributions of the element estimates in Section VI.A.

The modeling and estimation framework considered here assumes that the orbits are described by a set of constant parameters that do not vary in time. This paper mostly assumes the objects have Keplerian orbits, which implies that the elements or parameters describing each orbit are constant in time. Elements for other deterministic orbit models may vary over time, but the modeling and estimation framework adopted here are still applicable in these cases since these orbits can be parameterized with an initial position and velocity or element set. These initial conditions can then be treated as the constant parameters that are estimated for each orbit with EM; an example of this estimation is given in Section VII, where the Keplerian orbits are perturbed as a result of Earth being oblate. Other parameters characterizing the orbital motion, such as drag coefficients, could therefore also be estimated in principle by suitable expansion of the parameter space. The basic modeling framework assumed here does not extend to the case where process noise is used to account for perturbations as discussed further in Section VII.

As a motivating example, Fig. 1 shows four-hundred simulated azimuth and elevation angle measurements. The measurements are of four objects or clutter, and the observer is space-based with a Keplerian orbit as well; the supplement provides more details on how the data was simulated. Figure 1a shows the angle data without indicating data associations, so the measurements are not associated to specific objects or clutter. In Fig. 1b, the angle measurements

are associated to objects by color and clutter is marked as triangles. The corresponding time series plots of the angle measurements are shown in Fig. 1c-1f. Comparing the two sets of plots, it is not apparent which measurements belong to which objects or are clutter. Determining the orbits of the objects is required to reduce this uncertainty.

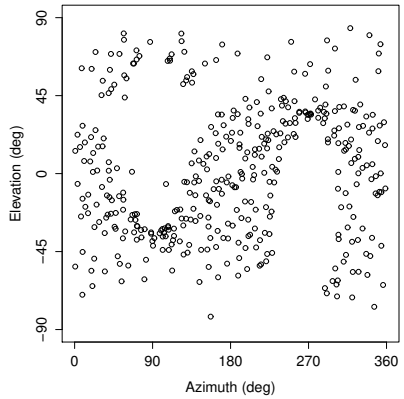
Clutter, or spurious measurements, may not be a major concern in practice for linking angle measurements of space objects, though it may be for other data types. This paper considers clutter to demonstrate the ability of EM to account for this phenomenon and as a possible model choice for unlinked measurements. Including a clutter model can help with orbit determination, even if no measurements are clutter, since it allows measurements to not be linked to a particular orbit. For example, suppose there are measurements of two objects, but both orbit models are inaccurate. If clutter is not included as a possible model choice and the measurements are initially associated to the incorrect orbit model, then it may be difficult to improve the orbit models. In contrast, if a clutter model is allowed, then the measurements can be associated as clutter until the orbit models are refined sufficiently so that the measurements can be linked correctly. Hence, clutter effectively acts as a temporary label for unlinked measurements until the measurements can be associated to the correct orbit model.

In order to determine the different orbits by estimating their elements, it is first necessary to specify a statistical model that is assumed to generate the data. Under certain assumptions, multiple object tracking data with unknown data associations can be modeled as a mixture model. The mixture model formulation of the tracking problem is described next and an EM algorithm for estimating parameters in the model, including the orbital elements, is discussed in Section IV. In general, different EM algorithms are required for different statistical models, but the procedure for constructing EM algorithms is standard for mixture models [55, 56].

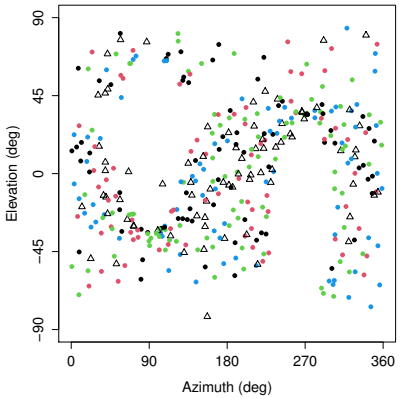
III. Statistical Model

This section combines the data association and measurement error models with the orbit models into a statistical framework called a mixture model. For notation, let J denote the number of orbit models to be included in the mixture model. Each object is assumed to have a single orbit model, so J is also the number of objects. The number of objects is often not known in practice but the multi-stage EM approach discussed in Section IV effectively allows this number to be estimated. Parameters in the mixture model that require estimation are denoted ξ and are listed at the end of this section.

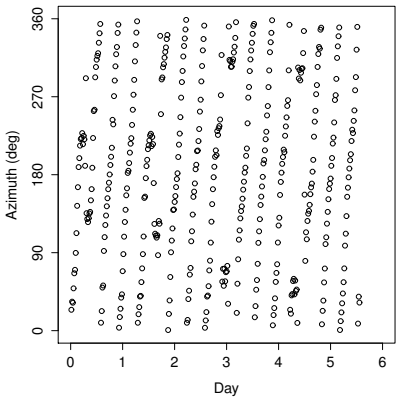
Since the data associations are not known, they are represented by an unobserved discrete-random variable whose values correspond to the different objects or clutter. In particular, the latent data association variable X_i is defined such that $X_i = 0$ if the ith measurement is of clutter and $X_i = j$ if the ith measurement is of the jth object for $j = 1, \ldots, J$. The ability of the observer to take a measurement of an object may depend on their separation distance, whether the line-of-sight from the observer to the object is blocked by Earth [57, 58], and other factors. However, for simplicity the



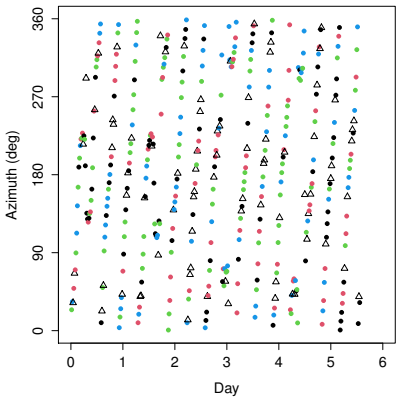
(a) Angle Data Without DA



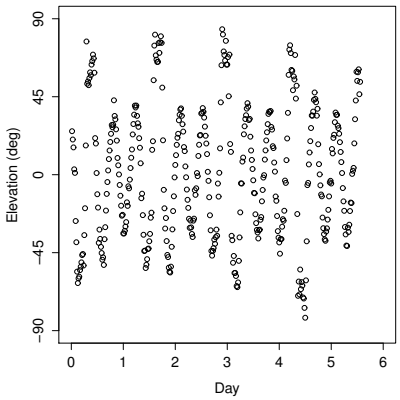
(b) Angle Data With DA



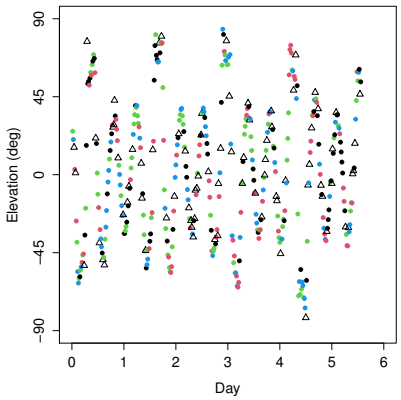
(c) Azimuth Data Without DA



(d) Azimuth Data With DA



(e) Elevation Data Without DA



(f) Elevation Data With DA

Fig. 1 Simulated data for four objects tracked in clutter. Data associations (DA) are either indicated (colors for objects, triangles for clutter) or not.

probabilities of observing clutter or one of the objects are modeled as constant in time,

$$p_{\xi}(X_i = j) = \pi_j, \tag{1}$$

for $j = 0, 1, \ldots, J$, which are non-negative and sum to one. The assumption that these probabilities are constant over the time the measurements are collected is reasonable if there is not significant prior information about the objects suggesting otherwise. The latent data association variables are assumed to be independent as well.

In general, the observer could be space-based or at a fixed-location on Earth's surface. The position of the observer is assumed to be known at all times and denoted $r_O(t)$. The position of the jth object at time t is denoted $r(t; \Theta_j)$, where Θ_j denotes the elements for object j. This paper parameterizes the Keplerian orbits with equinoctial elements [59], so that $\Theta_j = \{\tau_j, a_j, h_j, k_j, p_j, q_j\}$, where τ_j and a_j denote the time of perigee passage and semi-major axis, respectively. As described in [59], the remaining elements $\{h_j, k_j, p_j, q_j\}$ are functions of the eccentricity of the orbit and Keplerian elements defining the orbital plane; the supplement provides additional details and references on the relationship between equinoctial and Keplerian elements and an expression for the position of an object, $r(t; \Theta_j)$, with a Keplerian orbit. The time of perigee passage is included so that the elements can be treated as a set of fixed but unknown parameters requiring estimation. The relative position of the jth object with respect to the observer is

$$\tilde{r}(t; \Theta_j) = r(t; \Theta_j) - r_O(t), \tag{2}$$

which defines the mean direction along which angle measurements are obtained.

Angle measurements are assumed to be available at times t_i for $i = 1, \ldots, n$. The ith azimuth and elevation angle measurement is denoted $\phi_i = (\phi_{i,1}, \phi_{i,2})$ and the corresponding unit-vector measurement is denoted $Y_i = (Y_{i,1}, Y_{i,2}, Y_{i,3})$. Note that the ith unit-vector measurement can be obtained from the corresponding angle measurement via

$$Y_i = \begin{bmatrix} \cos(\phi_{i,2}) \cos(\phi_{i,1}) \\ \cos(\phi_{i,2}) \sin(\phi_{i,1}) \\ \sin(\phi_{i,2}) \end{bmatrix}, \tag{3}$$

and the reverse transformation from a unit vector to a pair of angle measurements is given by

$$\phi_{i,1} = \tan^{-1}(Y_{i,2}/Y_{i,1}) \tag{4}$$

$$\phi_{i,2} = \sin^{-1}(Y_{i,3}/\|Y_i\|). \tag{5}$$

For notation, Y_i denotes a random variable or realization of a unit vector, whereas y_i denotes the corresponding index

variable for the distribution of Y_i. Hence, if $p_\xi(y_i)$ is the distribution of the ith measurement Y_i, then this relationship is expressed as $Y_i \sim p_\xi(y_i)$, and $p_\xi(Y_i)$ is the contribution to the likelihood from Y_i. This notation extends to the discrete-random variable case, so $X_i \sim p_\xi(x_i)$ implies that the ith data association variable, X_i, has distribution $p_\xi(x_i)$, and $p_\xi(X_i)$ is the contribution to the likelihood from X_i.

Unit-vector measurements are assumed to follow the spherical normal distribution defined in [52], though other choices are possible [8, 60]. The spherical normal distribution is characterized by a mean-direction vector and a concentration parameter, λ, that controls the precision of the measurements. The mean-direction vector is taken to be the unit vector pointing at the object from the observer, or $\tilde{r}(t;\Theta_j)/\|\tilde{r}(t;\Theta_j)\|$. The spherical normal distribution of the unit vector associated with the ith angle measurement is

$$p_\xi(y_i|X_i = j) = [C(\lambda)]^{-1} \exp\left(-\frac{\lambda}{2} \arccos^2\left(\frac{y_i^T \tilde{r}(t;\Theta_j)}{\|\tilde{r}(t;\Theta_j)\|}\right)\right), \tag{6}$$

where the conditional dependency of the measurement on the jth object is made explicit and λ has units rad^{-2}. To interpret the distribution, note that $\arccos(u_1^T u_2)$ is the angular distance between two unit vectors u_1 and u_2. Hence, the spherical normal distribution is a Gaussian distribution over the angular distance measurement error. The normalizing constant $C(\lambda)$ is derived in [52] and given in equation 21 of their supplement. The constant is a non-linear function of the concentration parameter, but can be approximated as

$$C(\lambda) \approx \frac{2\pi}{\lambda} \tag{7}$$

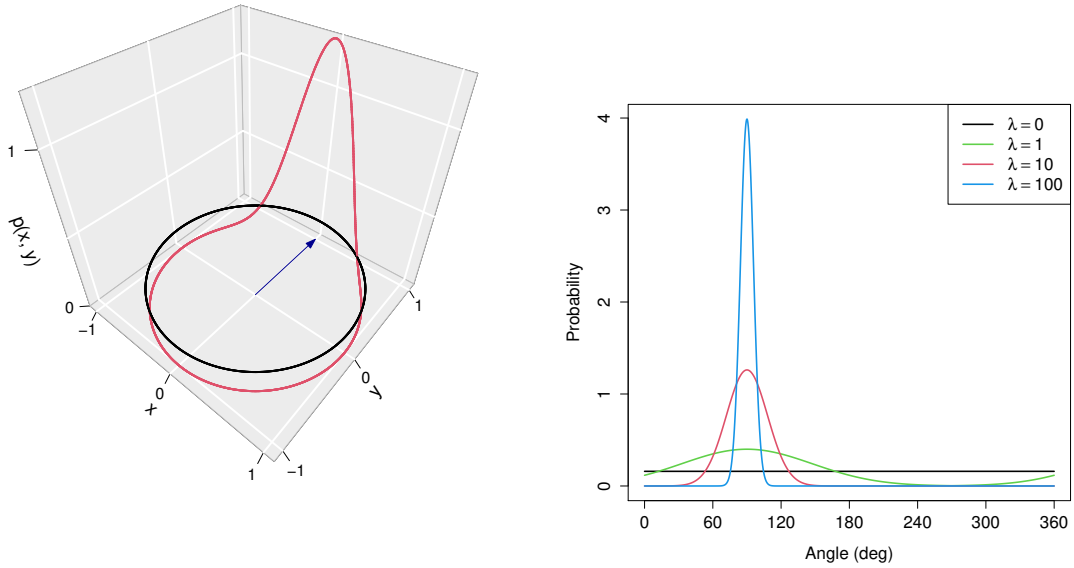
for large values of the concentration parameter, which is motivated by considering the normalizing constant for the bivariate normal distribution. The approximation error is negligible for the range of measurement errors considered here as illustrated in the supplement to this paper. Similar Gaussian approximations to a different spherical distribution are considered in [8].

As in [52], angle measurements of clutter are modeled as corresponding to unit vectors that are uniformly distributed over the unit sphere. Specifically, the clutter measurement distribution is

$$p_\xi(y_i|X_i = 0) = (4\pi)^{-1} \tag{8}$$

because the unit sphere has surface area 4π. More realistic clutter models are possible, such as uniform distributions over regions centered around objects or spatial Poisson processes [61], but this model is a reasonable default choice.

Figure 2a shows the two-dimensional spherical normal distribution for an object measurement and the uniform distribution for clutter. The origin represents the observer and the mean-direction vector, shown as an arrow, indicates



(a) Objects (red) and Clutter (black)

(b) Effect of Concentration Parameter

Fig. 2 Spherical normal distributions characterizing the error distribution for angles-only measurements.

the direction of the object. Note that the spherical normal distribution concentrates its probability in the direction of the object. In contrast, the clutter distribution is uniform on the unit circle, which reflects a lack of directional preference for clutter. Comparing the two distributions shows that an angle measurement aligned with the mean direction is more likely to be of an object than clutter, whereas an angle measurement orthogonal to the mean direction is more likely to be clutter than of an object. Figure 2b shows that the variance of the spherical normal distribution increases as the concentration parameter decreases. Note as well that the support of the probability distributions shown here is the unit circle, but the support for angles-only measurements is the unit sphere.

The statistical model described in this section is called a mixture model since the marginal distribution of a measurement is

$$p_\xi(y_i) = \sum_{j=0}^{J} p_\xi(y_i|X_i = j)\pi_j, \tag{9}$$

which follows by conditioning on the ith data association random variable. Note that the measurements, $Y_{1:n}$, are independent since the data association variables $X_{1:n}$ are independent; the notation $a_{i:j} = \{a_i, a_{i+1}, \ldots, a_{j-1}, a_j\}$ is used throughout this paper for any sequence a and indices $i \leq j$. In the terminology of mixture models, the π_j and $p_\xi(y_i|X_i = j)$ are called mixing probabilities and mixture components, respectively. Data association probabilities are defined as $p_\xi(X_i = j|Y_i)$, but discussion of how these probabilities are computed is deferred to the following section. The full mixture model is parameterized by the mixing probabilities, orbital elements, and concentration parameter, so

that $\xi = \{\pi_{0:J}, \Theta_{1:J}, \lambda\}$. The next section gives the EM algorithm for computing maximum likelihood estimates of the mixture model parameters and data association probabilities.

IV. Expectation-Maximization (EM)

EM is employed here to efficiently compute maximum likelihood estimates of the mixture model parameters, ξ, defined in the previous section. In principle, the likelihood could be maximized directly without EM, but this is difficult in practice since the data associations are not available. To see this, note that a maximum likelihood estimate (MLE) is a maximizer of the marginal likelihood function

$$L_Y(\xi) = \prod_{i=1}^{n} p_\xi(Y_i), \tag{10}$$

which is treated as a function of ξ and can be expressed as a product since the Y_i are independent. A log transformation is applied to the likelihood to establish the EM algorithm. The log-likelihood, $\ell_Y(\xi) = \log(L_Y(\xi))$, is

$$\ell_Y(\xi) = \sum_{i=1}^{n} \log(p_\xi(Y_i)) \tag{11}$$

$$= \sum_{i=1}^{n} \log\left(\sum_{j=0}^{J} p_\xi(Y_i|X_i = j)\pi_j\right). \tag{12}$$

The log-likelihood is difficult to maximize directly since it requires considering all possible data associations simultaneously. EM algorithms avoid this problem by separating the data association and parameter estimation problems, and therefore are often more effective at computing MLEs than methods that do not make this separation. In particular, the algorithms iterate between updating data association probabilities and parameter estimates, so that at each iteration data association probabilities are updated given the current parameter estimates, and the parameter estimates are updated given the current data association probabilities. EM algorithms are often used for parameter estimation when there is a latent variable that facilitates inference [62, 63], and are popular for fitting mixture models since the data associations are not known [52, 55, 64, 65].

EM algorithms are built on the joint log-likelihood of the measurements and the latent data association variables. The so-called complete-data log-likelihood (CDLL) for the mixture model is

$$\ell_{X,Y}(\xi) = \log\left(p_\xi(X_{1:n}, Y_{1:n})\right) \tag{13}$$

$$= \sum_{i=1}^{n} \left[\log(p_\xi(Y_i|X_i)) + \log(p_\xi(X_i))\right], \tag{14}$$

where $p_\xi(Y_i|X_i)$ is given in (6) and (8) for the different values X_i can assume and $p_\xi(X_i) = \pi_{X_i}$. Note, however, that

this function cannot be evaluated since the latent data association variables are not known.

An EM algorithm consists of iteratively taking the conditional expectation of the CDLL with respect to the measurements and maximizing the resulting function in the unknown parameters; these are called the expectation (E) and maximization (M) steps, respectively. The E step weights terms in the CDLL with data association probabilities computed from current parameter estimates, and the M step maximizes a related objective function with respect to the parameters to be estimated. These probabilistic data association and parameter update steps are then iterated until a measure of convergence is achieved. More formally, if $\xi^{(h)} = \{\pi_{0:J}^{(h)}, \Theta_{1:J}^{(h)}, \lambda^{(h)}\}$ denotes the mixture model parameter estimates of ξ at iteration h, then it can be shown that the parameter iterates $\xi^{(h)}$ converge to a local maximum or stationary point of the log-likelihood function (11) as the number of iterations, h, goes to infinity [62, 66]. Furthermore, for this implementation of EM, the estimated log-likelihood $\ell_Y(\xi^{(h)})$ does not decrease between iterations and tends to increase until $\xi^{(h)}$ converges. Hence, the algorithm is typically stopped when the estimated log-likelihood, $\ell_Y(\xi^{(h)})$, or the parameter estimates converge and the last $\xi^{(h)}$ is taken to be the MLE.

The EM algorithm for this mixture model consists of the following steps, which are provided here with explanatory details:

1) Initialization: Specify starting values for the parameters, $\xi^{(0)}$. Orbital elements can be initialized, for example, using an initial orbit determination method or sampling a range of plausible values; the latter approach is taken for the simulations in this paper based on hypothetical prior knowledge of the orbits. The mixing probabilities are initialized to be equal, so that $\pi_j^{(0)} = (J+1)^{-1}$ for all j, and the initial concentration parameter is taken to be $\lambda^{(0)} = 10^2$ rad^{-2}. Setting the initial estimate of the concentration parameter to reflect a large measurement error variance allows the orbital elements to move from their initial estimates to areas of the element space with higher likelihood. Set $h = 1$.
2) E step: Given the current parameter estimates, $\xi^{(h-1)}$, form the conditional expectation of the CDLL with respect to the measurements,

$$Q(\xi, \xi^{(h-1)}) = E_{\xi^{(h-1)}}[\ell_{X,Y}(\xi)|Y_{1:n}] \tag{15}$$

$$= \sum_{i=1}^{n} \sum_{j=0}^{J} \left[\log\left(p_\xi(Y_i|X_i = j)\right) + \log(\pi_j)\right] w_{ij}^{(h-1)}, \tag{16}$$

where the last term is the data association probability for measurement i and component j. The data association probabilities are computed using Bayes' rule as

$$w_{ij}^{(h-1)} = p_{\xi^{(h-1)}}(X_i = j|Y_{1:n}) \tag{17}$$

$$\propto p_{\xi^{(h-1)}}(Y_i|X_i = j)\pi_j^{(h-1)}, \tag{18}$$

where the normalizing constant is $p_{\xi^{(h-1)}}(Y_i)$ from (9) and measurements not from the ith time can be ignored since X_i and $Y_{i'}$ are independent for $i \neq i'$. In words, Bayes' rule updates the prior probability of observing component j, $\pi_j^{(h-1)}$, with the corresponding likelihood, $p_{\xi^{(h-1)}}(Y_i|X_i = j)$. Note from (6) that $Q(\xi, \xi^{(h-1)})$ is a weighted sum-of-squares function in the parameters ξ, where the $w_{ij}^{(h-1)}$ are the weights.

3) M step: Maximize $Q(\xi, \xi^{(h-1)})$ with respect to ξ to yield updated parameter estimates $\xi^{(h)}$. The updates for a set of orbital elements, $\Theta_j^{(h)}$, are not in closed-form due to the non-linearity of the log-likelihood, and so are obtained with a non-linear weighted least-squares solver called the Levenberg-Marquardt algorithm [67]. The weights are the data association probabilities, $w_{ij}^{(h-1)}$. Hence, data association uncertainty quantified in the E step is propagated into the parameter updates in the M step. The updates for the mixing probabilities can be computed in closed-form as averages of the data association probabilities from the E step,

$$\pi_j^{(h)} = \frac{1}{n} \sum_{i=1}^{n} w_{ij}^{(h-1)}, \tag{19}$$

which can be derived using a Lagrange multiplier argument since they sum to one [68]. Using the approximation (7), the concentration parameter can be updated as

$$\lambda^{(h)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{J} w_{ij}^{(h-1)}}{\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{J} \arccos^2 \left(\frac{Y_i^T \tilde{r}(t_i;\Theta_j^{(h)})}{\|\tilde{r}(t_i;\Theta_j^{(h)})\|} \right) w_{ij}^{(h-1)}}. \tag{20}$$

Updating the concentration parameter by maximizing $Q(\xi, \xi^{(h-1)})$ with respect to λ is more difficult without the approximation because of the form of the normalizing constant, $C(\lambda)$.

4) If convergence of the parameter estimates, $\xi^{(h)}$, or marginal log-likelihood, $\ell_Y(\xi^{(h)})$, is achieved or a pre-specified number of iterations is reached, stop and return $\xi^{(h)}$. Else, let $h = h + 1$ and return to step 2.

The EM algorithm should be run from multiple initial starting values to reduce the chances that a local maximum or stationary point of the likelihood surface is reached instead of the global maximum. Additional implementation details are provided in the appendix and supplementary material.

The E step for this EM algorithm reduces to computing the probabilities for each possible association of object to measurement. In particular, the association of measurement i with object j is the event $\{X_i = j\}$, and the probability of this association, conditional on the measurements, is $w_{ij}^{(h)}$. Each E step requires $n(J + 1)$ data association probabilities to be calculated, which is linear in the number of measurements, n, for a fixed number of orbit models, J. The parameter updates in the M step are made with these probabilistic associations, rather than binary data associations based on the most likely or most probable association. The E and M steps are repeated until the parameter estimates converge, so the total number of data associations requiring evaluation is $n(J + 1)$ multiplied by the number of iterations. Other EM

algorithms or methods for multiple object tracking may scale exponentially due to different modeling assumptions, as discussed further in Section VII.

In practice, the number of objects present in a set of measurements may not be known and must be estimated along with the elements for each of their orbits. There are several approaches to estimating the number of components in a mixture model, but they may require fitting the model with different numbers of components or specifying a large number of components. Additionally, if the number of orbits is large, then the estimation problem may be expensive or difficult since the parameter space is high-dimensional. However, one or more orbits may be identified in cycling through the E and M steps, even if an EM algorithm converges to a local maxima or all of the orbits in the mixture model are not identified. This suggests a multi-stage extension of EM in which measurements associated to orbits identified in one implementation of EM are removed from the original dataset before repeating the E and M steps on the remaining measurements.

The multi-stage EM procedure consists of appending the following steps onto the basic EM algorithm after convergence has been reached in step 4:

5) Remove from consideration any measurements that can be unambiguously linked to a particular orbit in the mixture model. If there are insufficient measurements to determine a new orbit or a different stopping criteria has been met, the E and M iterations can be ceased.
6) Otherwise, return to step 1 with the subset of measurements not linked to an orbit.

Each cycling through of the E and M steps is referred to as a stage, so that the standard EM approach where EM is always run on the full, original data set is referred to as single-stage EM. In contrast, in multi-stage EM the first cyling through of the E and M steps is the first stage, the second cycling through is the second stage, and so on. As the stages progress, the number of measurements is reduced as some are linked to specific orbits, which then simplifies the estimation problem in subsequent stages since there are fewer measurements. Multi-stage EM may not find a local maximum of the log-likelihood for all of the measurements, but is still guaranteed to converge to a local maximum of the likelihood function at each stage. Also, the estimated mixing probabilities and data association probabilities from multi-stage EM are less interpretable than in single-stage EM since they are relative to the data available at each stage. Despite these potential drawbacks, multi-stage EM allows a large data set to be fit in smaller pieces and so enables scaling of EM to larger data sets than standard EM.

This section is concluded with a summary of EM and multi-stage EM. At each EM iteration, data association probabilities are computed in the E step given the current parameter estimates. The probabilities combine information from the data, through the likelihood function, with prior information about the data association. The M step updates estimates of the orbital elements by solving a weighted least-squares optimization problem, where the weights are the data association probabilities from the E step. Iterating between the probabilistic data association and parameter update steps produces an MLE of the mixture model parameters. In multi-stage EM, measurements associated to specific

objects are removed from consideration and the E and M steps are repeated to maximize the likelihood of the remaining measurements. Single-stage and multi-stage EM are demonstrated on simulated data in the following section.

V. Tracking Multiple Objects in Clutter

This section presents results for fitting data generated from the tracking scenario considered thus far, where an angle measurement of one of multiple objects or clutter is sampled at each time. In Section V.A, the single-stage EM algorithm comprised of steps 1-4 in Section IV is applied to the data shown in Fig. 1. In Section V.B, multi-stage EM is shown to identify more orbits than single-stage EM for a data set with ten orbits and is shown to scale-well as the number of orbits is increased.

A. Single-Stage EM

Single-stage EM is applied to the simulated data shown in Fig. 1 for six different initializations of the orbital elements. Initial orbital-element estimates, $\Theta_{1:J}^{(0)}$, are sampled uniformly from a range containing the true values of the elements. The initial estimates of the mixing probabilities and concentration parameter are given in Step 1 of the EM algorithm discussed in Section IV. Figure 3a shows the estimated log-likelihoods, $\ell_Y(\xi^{(h)})$, for all six runs, and the log-likelihood at the true parameter values, $\ell_Y(\xi)$. Note that the likelihoods do not decrease, and often increase, at each iteration as discussed in Section IV, and that all of the runs converge within the maximum number of iterations. The estimated log-likelihoods for five of the initializations do not converge to the largest likelihood value, suggesting they converged to a local maxima or stationary point of the log-likelihood surface instead. The estimated log-likelihood for the fourth initialization, referred to as run four, converges to the actual log-likelihood, $\ell_Y(\xi)$, and attains the largest likelihood overall. Hence, the parameter estimates for run four are selected as the maximum likelihood estimate from the ensemble of runs.

Data association accuracy is quantified with the Rand index [54], which in this context gives the percentage of pairs of measurements that are correctly linked to the same orbit or correctly linked to different orbits. The Rand index values for the six runs are shown in Fig. 3b to increase across EM iterations, indicating improved data association accuracy as parameter estimates improve. Furthermore, the log-likelihood and Rand index values in Fig. 3a and Fig. 3b increase together since better data assocations lead to better parameter estimates and vice-versa. For runs 3 and 4, the Rand index approaches one, indicating that all pairs of measurements have been correctly linked as belonging to the same or different orbits. Here, the ith measurement is linked to the orbit for which it has the largest data association probability; that is, the measurement is linked to the jth orbit, or more generally mixture model component if clutter is considered, if $w_{ij}^{(h)} \geq w_{ij'}^{(h)}$ for $j' = 0, \ldots, J$. The Rand index is therefore computed as

$$RI = \frac{n_1 + n_2}{\binom{n}{2}}, \tag{21}$$

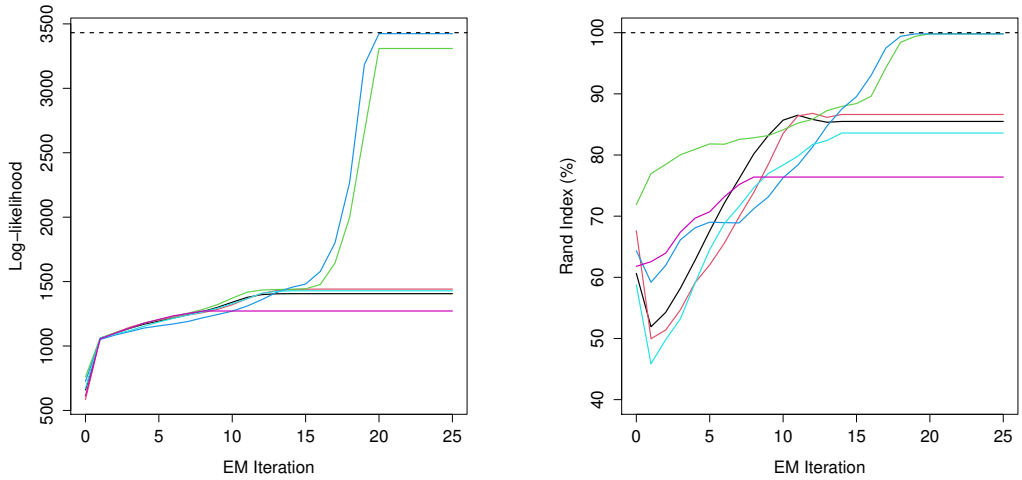
where n_1 is the number of pairs of measurements correctly linked to the same orbit, n_2 is the number of pairs of measurements correctly linked to different orbits, and n is the number of measurements. The denominator in the Rand index is a binomial coefficient giving the number of pairs of measurements and equals $n(n-1)/2$. Comparing data associations pairwise avoids label switching problems due to arbitary component labels, which can make assessing clustering algorithms difficult. [24] discusses similar statistics for quantifying data association accuracy, but a comparison of these statistics is left as future work.

Clutter identification is also investigated as a special case of data association. The ith measurement is identified as clutter if the clutter mixing probability is larger than all of the object data association probabilities, so that $w_{i0}^{(h)} > w_{ij}^{(h)}$ for $j = 1, \ldots, J$. Figures 3c and 3d show the accuracy and false discovery rate (FDR) of identifying clutter, respectively. Accuracy is the percentage of measurements that are correctly identified as clutter or not clutter, whereas the false discovery rate (FDR) for identifying clutter is the percentage of object measurements that are incorrectly identified as clutter. Hence, the accuracy plot shows that EM correctly differentiates clutter and object measurements. In contrast, the FDR plot shows that EM does not incorrectly identify object measurements as clutter. While an accuracy of one implies an FDR of zero, the converse is not true, so both plots may be helpful in diagnosing incorrect data associations. Comparing all of the plots in Fig. 3 indicates that identifying clutter improves data association for all mixture model components and vice-versa.

Figure 4a shows the convergence of the concentration parameter estimates to the actual value used in the simulation. The corresponding convergence of the standard deviations of the angular distance measurement errors is shown in Figure 4b; note that an expression for the variance, and equivalently the standard deviation, as a function of the concentration parameter is given in [52]. The estimates for the different EM runs are in sync with the data association results in Fig. 3. Specifically, improvements in data association accuracy occur as the measurement error variance decreases, and the runs that result in a close estimate of the concentration parameter are also the runs that achieve high data association accuracy.

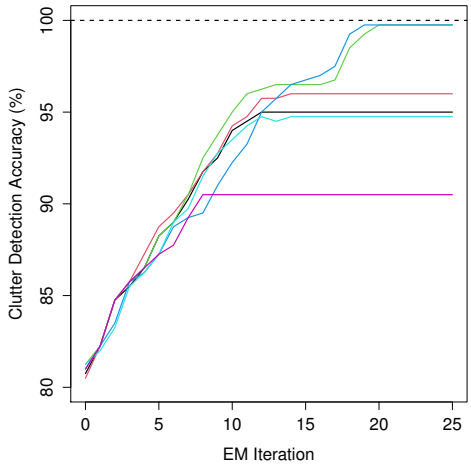
Figure 5 shows estimates of the orbital elements across EM iterations for the run that led to the largest final estimated log-likelihood. The convergence of the estimates from their initial values to the true values is apparent for all four sets of orbital elements, demonstrating that the EM algorithm can correctly determine the orbits of the objects. Figure 6 shows the mixing probability estimates simultaneously converging to the actual values. In the first iteration, the mixing probabilities for the objects that are most consistent with the measurements increase sharply and the other mixing probabilities decrease towards zero since the probabilities sum to one. As the iterations progress and the orbital-element estimates converge to the true values, the mixing probabilities converge to their actual values. Note as well that the time-independent observation probability assumption in (1) does not appear to have a biasing effect on estimates of the orbital elements or mixing probabilities.

The supplementary material discusses diagnostic plots for validating the orbit model estimates based on prediction

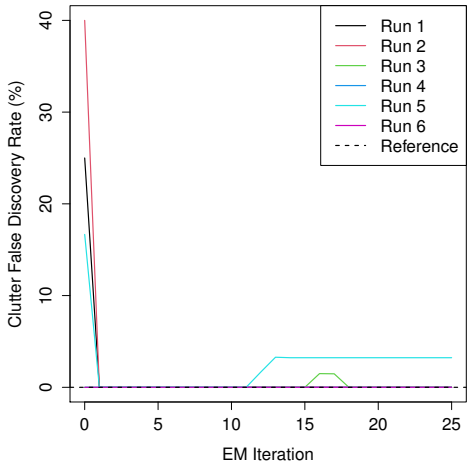


(a) Log-likelihood

(b) Rand Index

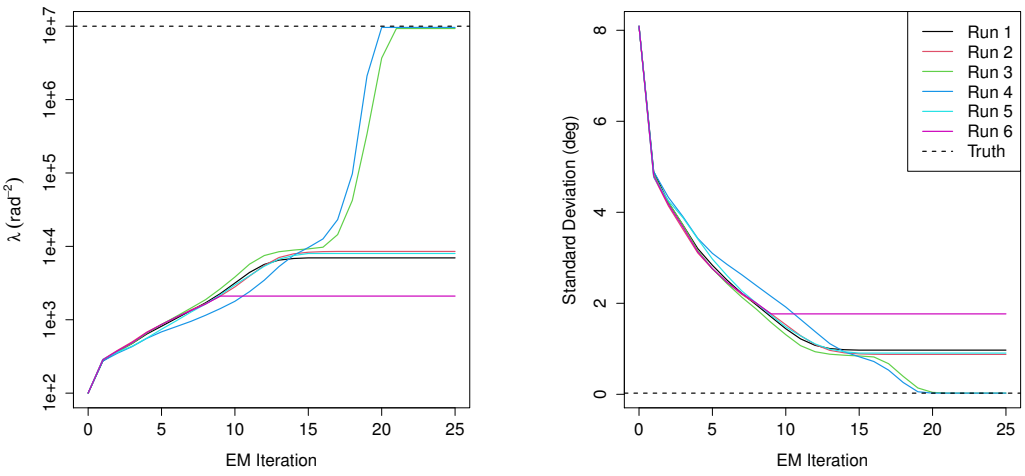


(c) Clutter Detection Accuracy



(d) Clutter False Discovery Rate

Fig. 3 Figure 3a shows the estimated log-likelihoods and actual log-likelihood (dashed line). Figures 3b-3d show data assocation statistics and reference values (dashed lines). Colors indicate different EM initializations.



(a) Concentration Parameter

(b) Standard Deviation

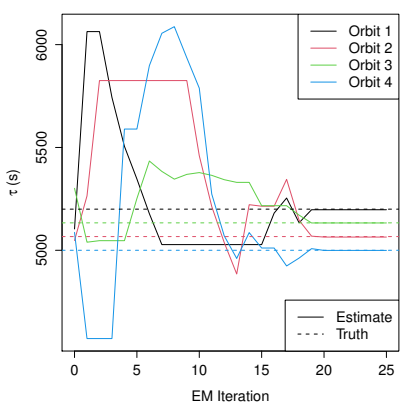
Fig. 4 Estimates of the concentration parameter and the standard deviation of the measurement errors for different EM initializations. The dashed black lines indicate the actual values.

residuals. In particular, time series and quantile-quantile plots are suggested for assessing the independent and identically distributed measurement error assumptions.

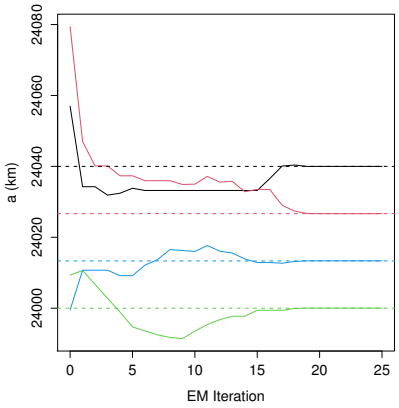
The results in this section demonstrate that EM is effective at orbit determination and probabilistic data association for a particular data set when the number of objects is assumed known. In practice, however, the actual number of objects may not be known. Furthermore, EM will produce an estimate of elements for J orbit models, but some of these orbits may not be accepted as real if they are inconsistent with the measurements. That is, the number of orbits accepted at the maximum likelihood estimate as actual orbits may be less than J or the actual number of objects. This may be indicative of EM reaching a local maxima of the likelihood function or misspecifying J. The following section shows that multi-stage EM can use this information to determine a larger number of orbits than single-stage EM and considers an example where the number of objects is unknown.

B. Multi-Stage EM

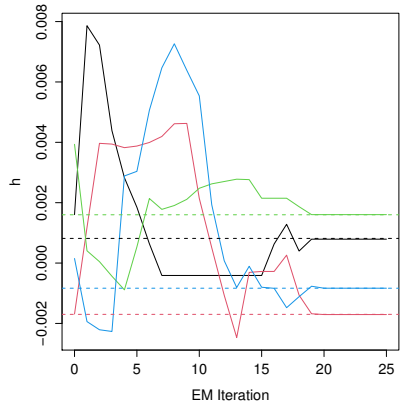
To motivate the need for multi-stage EM, fifty data sets are simulated as in the previous simulation and each fit with single-stage EM until at least one orbit is obtained. Each data set consists of one-thousand measurements from ten objects; clutter measurements are excluded from this simulation but a clutter model is included in the mixture model for reasons discussed in Section II. The orbital elements for the observer are the same in each data set, but are varied across objects as discussed in the supplement. Since the number of orbits is not assumed known, $J = \max(\lceil n/80 \rceil, 2)$ orbits are initialized, where n is the number of measurements and $\lceil x \rceil$ is the ceiling function. Note that n will decrease across stages as orbits are identified and the associated measurements are removed from the data set. For this simulation, the



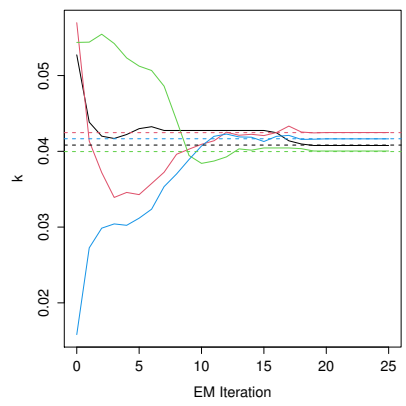
(a) Estimates of $\tau_{1:J}$



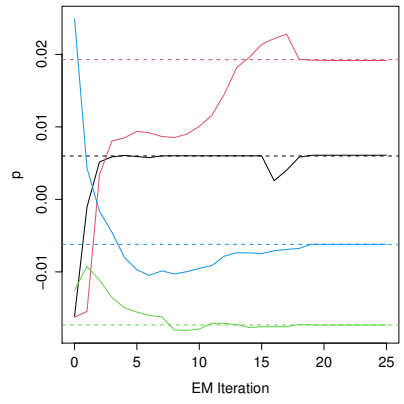
(b) Estimates of $a_{1:J}$



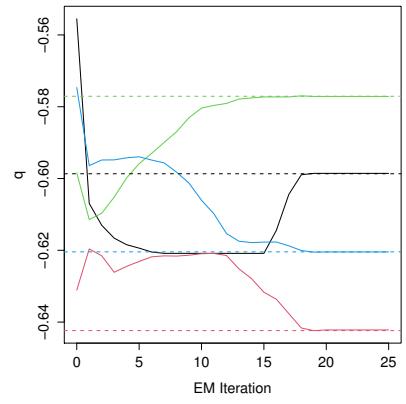
(c) Estimates of $h_{1:J}$



(d) Estimates of $k_{1:J}$



(e) Estimates of $p_{1:J}$



(f) Estimates of $q_{1:J}$

Fig. 5 Element estimates (solid lines) for run four and their actual values (dashed lines) across EM iterations. Each color represents a different orbit.

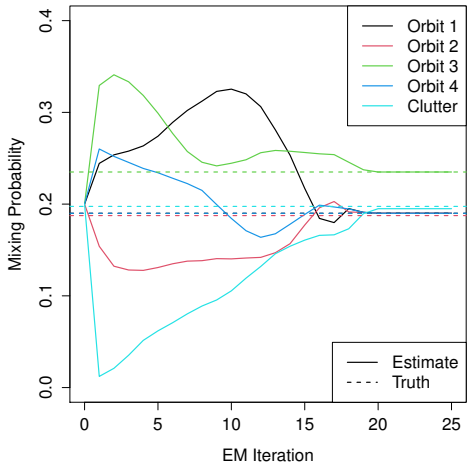
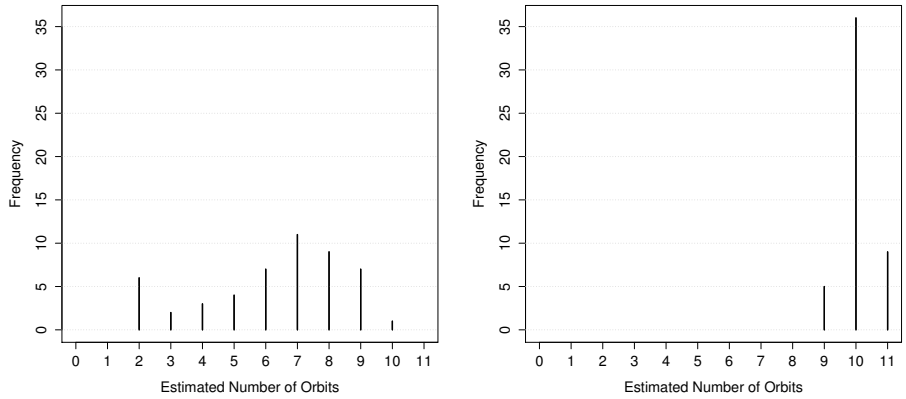


Fig. 6 Mixing probability estimates (solid lines) for run four and actual proportions in the data (dashed lines). Colors for orbit models match Fig. 5.

jth mixture component is declared to be an actual orbit if more than 40 measurements have $w_{ij}^{(h)}$ as their largest data association probability and if the maximum angular distance residual for that set of measurements and orbit is 250 arcseconds, which is approximately three standard deviations of the measurement error distribution.

Figure 7a shows the estimate of the number of orbits obtained with single-stage EM for the fifty data sets. For all but one of the data sets, single-stage EM identifies less than the total number of orbits due, for example, to EM converging to a point in the parameter space corresponding to a subset of the total number of orbits. One option would be to run single-stage EM for each simulated data set multiple times and vary the initial starting values each time as in Section V.A, but this can be computationally expensive and does not take advantage of information obtained in previous runs. Instead, we proceed with implementing multi-stage EM by removing measurements corresponding to any identified orbits and applying EM again to the remaining measurements. Figure 7b shows the estimate of the number of orbits obtained with multi-stage EM. Multi-stage EM does not always determine the correct number of orbits, but in general identifies more orbits than single-stage EM and in most cases identifies the correct number of orbits. Hence, this simulation illustrates that multi-stage EM can scale better than single-stage EM to linkage problems with tens of objects.

Figure 8 shows how many orbits are identified at each stage of multi-stage EM for three different initializations. For the first run, three orbits are identified in the first and fourth stages, two orbits are identified in the third stage, and one last orbit is identified in the tenth stage for nine orbits total. Similarly, for the second run, four orbits are identified in the first and third stages and one orbit is identified in the second and fourth stages for ten orbits total. Note that zero orbits may be identified in some stages, and all multi-stage EM runs shown in this figure converged by the tenth stage as determined by a pre-specified stopping criteria. The stopping criteria will vary by need, but here is taken to be if either less than fifty measurements are left unlinked or sixty stages have been completed.



(a) Single-stage EM (b) Multi-stage EM

Fig. 7 Estimated number of orbits with single-stage and multi-stage EM.

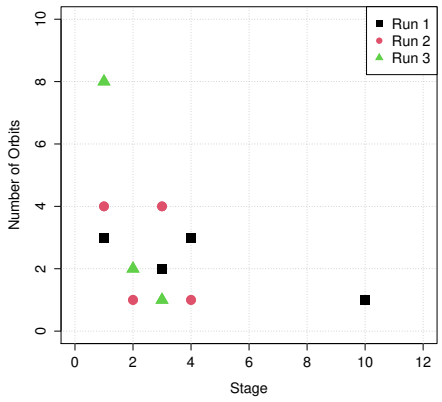


Fig. 8 The number of orbits identified across stages in three different initializations of multi-stage EM.

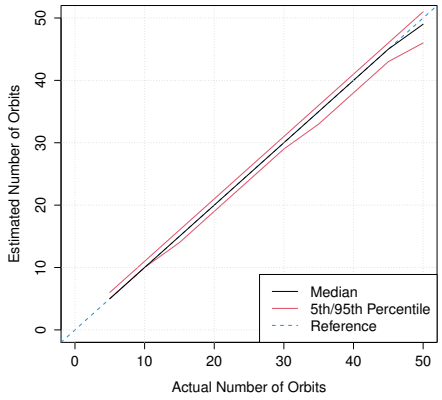


Fig. 9 Multi-stage EM estimates of the number of orbits (5th, 50th, and 95th percentiles) and a reference line in blue.

Figure 9 shows how multi-stage EM scales as the number of objects increases. As the number of orbits is increased from five to fifty in increments of five, one-hundred data sets are simulated and fit with multi-stage EM. The figure shows percentiles of the estimated number of orbits across the hundred data sets. In this simulation, multi-stage EM identifies the correct number of orbits in most of the data sets with a slight increase in variance and bias as the number of objects approaches fifty.

A consideration in multi-stage EM is determining when a measurement is linked to an orbit since linked measurements are not considered in later stages as possibly belonging to different orbits. The linking criteria used in this example was based on hypothetical knowledge of the simulation settings in this paper; other criteria will be necessary for different orbit propagators or measurement error distributions. This consideration is not required for single-stage EM because measurements are not linked to specific orbits, but rather are given probabilities of belonging to the different orbits. Measurements of ambiguous origin can be identified after running multi-stage EM by computing data association probabilities under the orbit models identified across all stages. If there are measurements consistent with two or more orbit models, then EM could be applied to the measurements linked to these orbits to try to refine the associated orbit models and determine the association of the measurements whose origins are not clear. Similarly, the criteria to identify an orbit as real will depend on how many measurements are available for each object, the distribution of the measurement error, and the fidelity of the force propagator.

Another consideration with multi-stage EM is the number of orbit models to include in the mixture model at each stage, though this consideration is also present for single-stage EM. In multi-stage EM, if J_k denotes the number of orbit models included in the mixture model at stage k, then up to $\sum_k J_k$ orbit models can be determined overall. However, in single-stage EM the number of orbits that can be determined is at most J. Hence, multi-stage EM allows more flexibility in specifying the number of orbit models in the mixture model than single-stage EM.

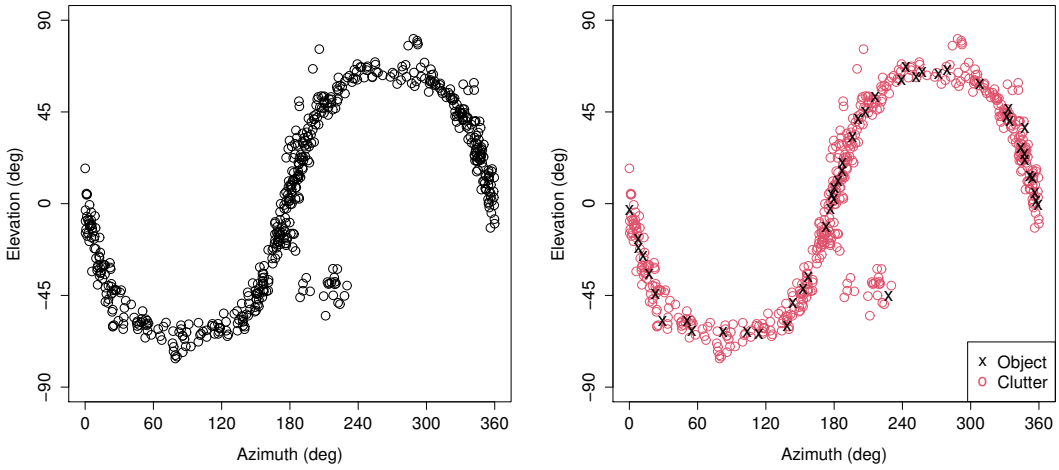
The next simulation considers single-stage EM again and quantifies the bias and uncertainty in EM estimates over multiple data sets. A different tracking scenario is assumed to demonstrate that EM is adaptable to different tracking scenarios.

VI. Tracking a Single Object in Clutter

An EM algorithm is now discussed for a second tracking scenario in which a single object is tracked in clutter, but multiple measurements of clutter are allowed at each measurement time. The detection model assumes that the object is observed with the same probability of detection at each time and that the object detections are independent. Hence, it is not known whether there is a detection at a particular time, and if there is a detection, the measurement index of the object is not known. The structure of this tracking scenario is therefore similar to the previous scenario with $J = 1$ objects and probabilities of detection and non-detection given by $p_D = \pi_1$ and $1 - p_D = \pi_0$, respectively, but now the detection model is slightly different due to the multiple clutter measurement possibility. The mixture model and EM algorithm for data generated from this scenario are given in the appendix. A more general model would combine the model in this section with the previously discussed model by allowing for multiple objects and measurements at each observation time, though the number of possible data associations to consider here would be large. Section VI.A discusses approximations of the distribution of the estimates of the elements.

For this tracking scenario, the detection probability, orbital elements of the object, and concentration parameter are not known and so are estimated with EM. As in the previous scenario, the EM algorithm for this model iterates between computing data association probabilities for object detections in the E step and updating estimates of the mixture model parameters in the M step. Previous work on tracking a single object considers the probabilistic data association (PDA) filter [11, 69], nearest neighbor filter [70], and EM or probabilistic MHT (PMHT) algorithms [42, 71, 72]. The PDA filter and PMHT approaches are compared in [42]. In general, when the motion of an object is described by a parametric set of equations, estimating the motion model parameters by directly maximizing the likelihood may be feasible when there are few parameters to estimate. However, this approach generally becomes less efficient than EM at parameter estimation as the number of parameters increases.

To demonstrate the performance of the EM algorithm over a range of conditions, one-thousand data sets are simulated at detection probabilities of $i/10$ for $i = 1, \ldots, 10$. Each dataset consists of measurements collected at one-hundred times, where the semi-major axis and eccentricity elements are the same in each data set but the other orbital elements are randomly varied across data sets. At each time the number of clutter measurements is simulated from a Poisson distribution with mean five. Angle measurements of clutter are simulated uniformly around the angle of the object, so the clutter tends to follow the object as it moves through space. Furthermore, at each time the random variable representing an object detection follows a Bernoulli distribution with the detection probability as the probability of success. An example data set is shown in Fig. 10, where the detection probability is 0.5. For each simulated data set,



(a) Angle Data Without Associations

(b) Angle Data With Associations

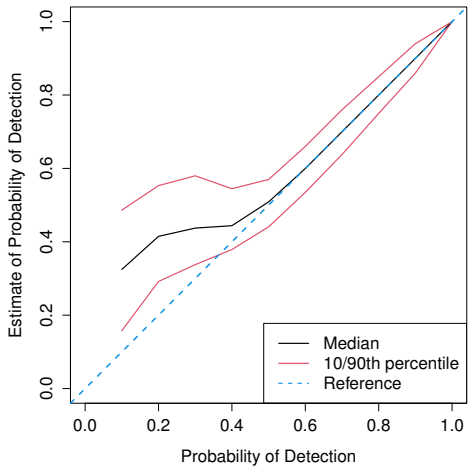
Fig. 10 Example data for Section VI. In Fig. 10a, the data associations are not known. Fig. 10b indicates the object or clutter associations.

the EM algorithm is run fifty times with different, randomly selected initial starting values of the orbital elements and detection probability. Of the fifty EM runs, the run with the largest final log-likelihood estimate is saved and the Rand index and parameter estimates are recorded. As for the previous tracking problem, the goal is to determine which measurements are of the object in orbit and to determine the orbital elements of the object.

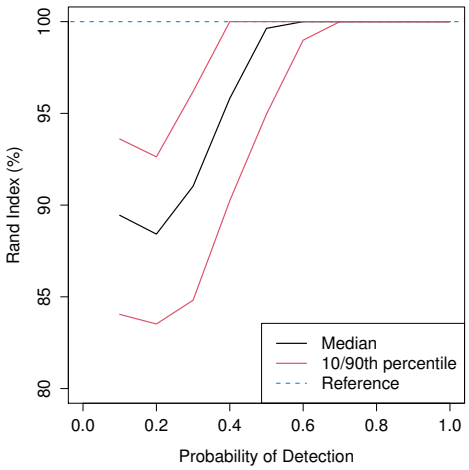
Figure 11 shows summary results of the Rand index, detection probablity, and concentration parameter estimates. Similarly, Fig. 12 shows results of the semi-major axis and eccentricity estimates; analogous plots are not shown for the other orbital elements since they were varied across simulated data sets. In general, the estimates at low detection probabilities have large variance and some estimates, such as the concentration parameter, appear to be biased. As the detection probability increases, the bias and variance is reduced since more measurements of the object are available to estimate the elements and other mixture model parameters. Figure 11b indicates that reduced variance in the parameter estimates is also associated with increased data association accuracy. Note as well the relatively small concentration parameter estimates in Fig. 11c at low detection probabilities. The measurement error model is poorly learned at low detection probabilities since there are insufficient object measurements to learn the orbit model well.

A. Distributions of Element Estimates

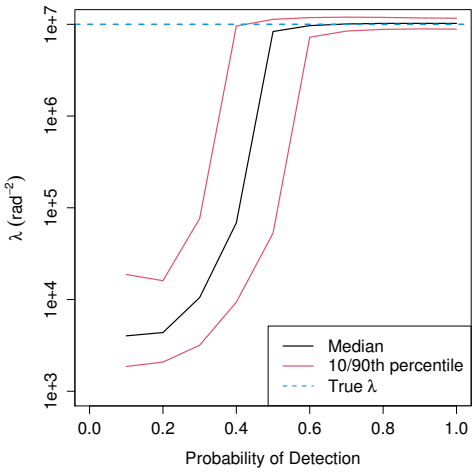
This paper uses EM to compute maximum likelihood estimates of orbital elements, and therefore methods for approximating the distribution of maximum likelihood estimates are applicable. Two common approaches for this task are to approximate the maximum likelihood estimates as Gaussian or to use bootstrapping; a detailed discussion and comparison of these approaches is provided in [63] or [56], for example. The former approach relies on approximating



(a) Detection Probability

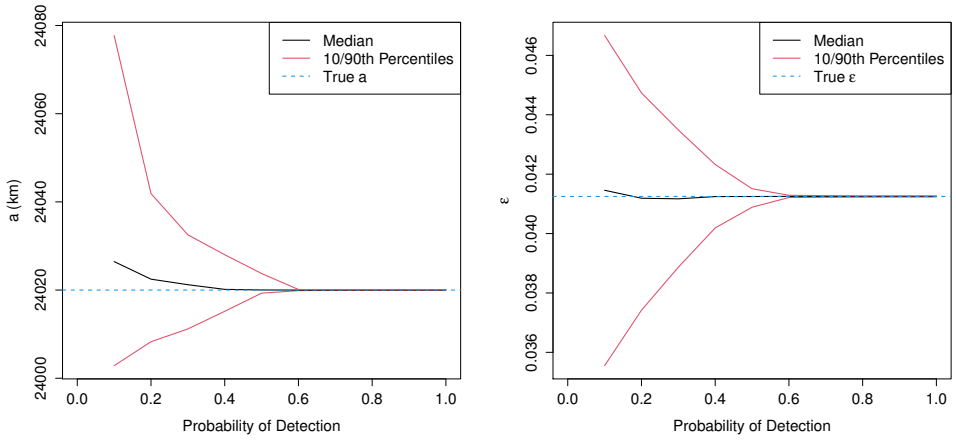


(b) Rand Index



(c) Concentration Parameter

Fig. 11 Estimate percentiles across one-thousand simulated data sets for tracking a single object in clutter.



(a) Semi-major Axis

(b) Eccentricity

Fig. 12 Element estimate percentiles (solid lines) and their actual values (dashed lines).

the asymptotic variance-covariance matrix of the element estimates, and so in principle requires a large number of object measurements. Bootstrapping, in contrast, requires simulating multiple data sets from the fitted mixture model and estimating the model parameters for each of the new data sets; Fig. 13 shows an example of element estimates from a bootstrap distribution for a simulated data set with $p_D = 0.4$. Confidence intervals for the elements can then be obtained from the sample covariance matrix or quantiles of the bootstrapped estimates. Overall, the Gaussian approximation is less expensive to implement than bootstrapping for this application, but is more difficult to justify in a small data setting.

Figure 14 shows the confidence interval coverage for the six orbital elements using a Gaussian approximation and bootstrap method. The confidence interval coverage is the percentage of confidence intervals that contain the actual values of the elements, which are known because the data is simulated. Here, the confidence intervals are constructed at a nominal confidence level of 95%, so their coverage should be approximately the same over multiple data sets. In Fig. 14a, the coverage of the intervals from the Gaussian approximation increases with the detection probability. This trend is expected since the Gaussian approximation improves as the number of measurements increases, though the coverage is high for values of p_D above 0.6.

Figure 14b shows the bootstrap confidence interval coverage, which is close to the nominal level of 95% for all values of p_D for the p and q elements. The coverage for the elements a and k increases from about 60-70% at $p_D = 0.1$ to 100% for $p_D = 0.6$, and the coverage for the elements τ and h is close to 100% for all values of p_D. The near 100% coverage for both confidence interval types may be due to the small measurement errors and the large number of measurements, which allows the elements to be estimated with high-precision and accuracy. In contrast, the less-than-nominal coverage for small values of p_D is likely due to inaccurate estimates of the elements resulting from a small number of measurements of the object. Gaussian confidence intervals are centered around the element

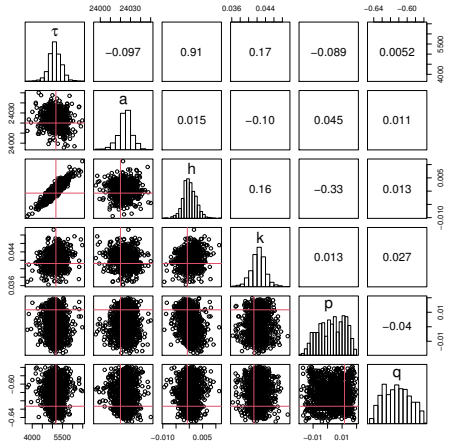
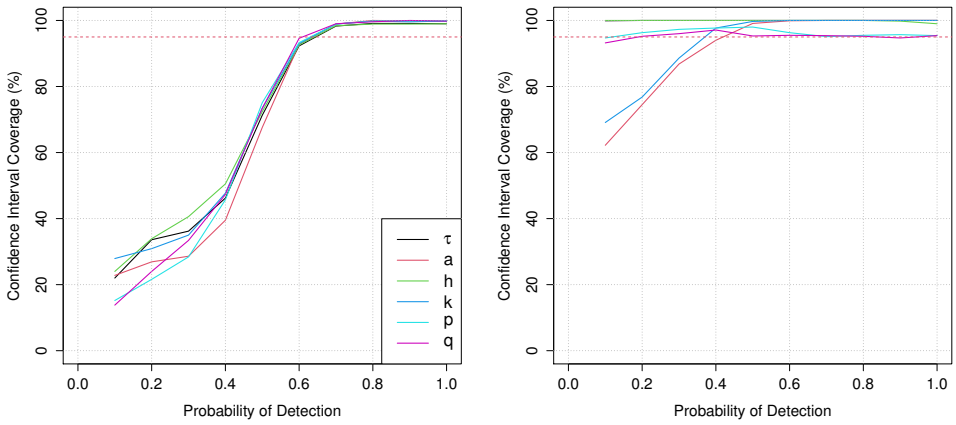


Fig. 13 Example distribution of bootstrapped orbital-element estimates (black circles) and maximum likelihood estimates (red lines). The numbers are the sample correlations of the bootstrap estimates.

estimates, so if the element estimates and their standard errors are inaccurate, then the confidence intervals are not expected to have the nominal coverage. Additional sources of variability in the coverage include Monte Carlo error due to simulation of data sets. Improving the coverage of bootstrap confidence intervals by, for example, considering alternative procedures for constructing the bootstrap distribution is left as future work.

The assumed mixture model does not lead to quantifying state uncertainty with a filter, though data association uncertainty is accounted for when updating elements in the M step. Filter-based EM algorithms for multiple object tracking are considered in [34] or [46], but may not be necessary if process noise can be ignored. When process noise can be ignored, EM produces point estimates of orbital elements by processing all measurements simultaneously. Alternatively, a Bayesian method could be employed to quantify uncertainty in the elements with a posterior distribution, which would lead to a different form of data association uncertainty than the weights computed with EM, but this is often more expensive than computing maximum likelihood estimates with EM since the joint parameter and data association space are high-dimensional. The trade-off between the quality of these approximations and computational cost should be considered when choosing a method for analysis in practice. In particular, fitting a mixture model with EM is often cheaper than implementing Bayesian inference of a state space model, but the latter quantifies uncertainty more naturally.

Summary statistics across ensembles of simulated data sets provide insight into how EM may perform for a new data set. This simulation estimated the confidence interval coverage for orbital-element estimates as the detection probability changes, for example. For the simulation settings considered here, EM tends to provide accurate estimates even at low detection probabilities. However, the performance is expected to deteriorate if less data are collected or clutter is more prevalent. Additional limitations of EM and modeling issues are discussed in the following section.



(a) Asymptotic Gaussian Approximation

(b) Bootstrap

Fig. 14 Empirical confidence interval coverage of orbital elements. The red dashed line is at 95%.

VII. Discussion

This paper considers the use of multi-stage EM for the multiple space object tracking problem when the number of objects is not known. Recall that the procedure is to specify a number of objects, fit the corresponding mixture model with EM, remove measurements corresponding to any identified orbits, and repeat these steps with the same or smaller number of orbit models in the mixture model until a stopping criteria is met. Alternatively, the number of orbit models can be set to an upper limit on the number of objects believed to be present. The hope is then that there are only as many non-zero mixing probabilities as there are objects. In practice, however, multiple orbit models can describe the same object, so regularization is often used to keep the orbit models distinct. A standard approach to allowing a large number of mixture model components that incorporates this regularization is with a Dirichlet process; details on its incorporation into EM algorithms are available in [73]. The general idea is to weight the mixing probabilities towards zero, so that most mixing probabilities are effectively zero. Alternative approaches for tracking an unspecified number of space objects are given in Section I.

The computational cost of a multiple object tracking algorithm depends on the assumed statistical model, the force propagator, and other choices. For maximum likelihood estimation, the cost is therefore in large part dependent on the likelihood of the model and how efficiently the likelihood can be maximized. Recall that the log-likelihood for the mixture model consists of $n(J+1)$ terms by equation (11), which increases linearly in the number of orbit models for a fixed number of measurements. Directly maximizing the likelihood of a mixture model is generally intractable, but EM efficiently computes maximum likelihood estimates by iteratively maximizing the objective function defined in (15). Other procedures for computing maximum likelihood estimates of a mixture model will have the same cost of evaluating the likelihood, but will generally require more likelihood evaluations and so are less efficient and more expensive overall.

In a state space model formulation of the multiple object tracking problem, the number of terms in the likelihood generally increases faster than linearly due to the dependency structure of the model; recall that state space models generalize mixture models to allow for dependency in the latent state or data association processes. Specifically, the number of terms can increase exponentially with the number of measurements, so that the likelihood cannot always be evaluated in practice. Some EM or EM-based algorithms for state space models scale linearly [36, 46], but others scale exponentially [34]. In the latter case, a track management procedure such as pruning is required to remove unlikely data associations from consideration. This discussion is meant to emphasize that the complexity of an EM algorithm for multiple object tracking depends on the assumed underlying model and is not inherent to EM itself. The advantage of using EM for multiple object tracking is not necessarily its scaling, which varies by model and implementation, but the framework it provides for probabilistic data association and efficient maximum likelihood estimation.

The main computational cost of fitting a mixture model with EM is updating the elements in the M step since it requires evaluating each orbit model multiple times as part of a non-linear least squares solve. As with other methods, the cost of updating the elements therefore increases with the fidelity of the orbit model, the number of measurements, and other possible factors. There are several approaches to mitigating the cost of EM, if needed, but two are discussed here, namely generalized EM and increasing the fidelity of the orbit model across EM iterations.

Generalized EM (GEM) reduces the computational cost of EM by effectively replacing the maximization, or M, step with a step that updates the elements by increasing the objective function; details are provided in [62] or [56], for example. Increasing the objective function can be more computationally tractable than finding a local maximum and also move the elements towards areas of the element space with higher likelihood. As the iterations progress, more computational resources can be made available to implement a full maximization in the M step. To demonstrate the advantages of GEM, Fig. 15 compares the run times and Rand indices for fifty initializations of single-stage EM, multi-stage EM, and multi-stage GEM. The simulated data set for this example consists of one-thousand measurements of ten objects or clutter. The element-update step in the GEM algorithm is obtained by limiting the number of Levenberg-Marquardt (LM) iterations at each M step, so there is no guarantee that the updated elements will maximize the objective function in the M step. Specifically, the number of allowed LM iterations starts at two and is increased to five, ten, and twenty at the tenth, twentieth, and thirtieth EM iterations, respectively. In general, Fig. 15 shows that single-stage EM is the slowest method and has the smallest data association accurary on average as judged by the Rand index statistic. Multi-stage EM is faster than single-stage EM and tends to have better data association accuracy. Multi-stage GEM is the fastest method of the three considered methods and has the highest average Rand index value, suggesting GEM has additional data association accuracy benefits. The improvement in data association accuracy may be due to less aggressive updating of the elements initially, which could lead to finding local maxima of the likelihood function. Further investigation is needed to investigate the rate at which the number of LM iterations should be increased across EM iterations.

A second approach to mitigating the cost of EM is to increase the fidelity of the orbit model over successive EM

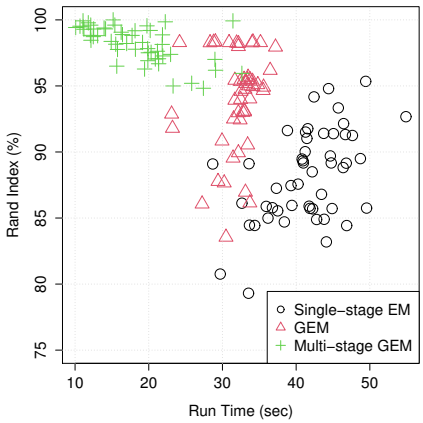
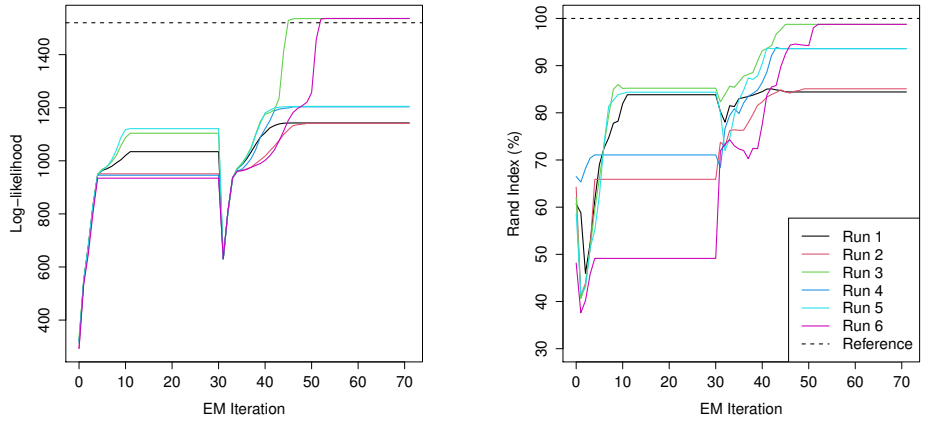


Fig. 15 Comparison of single-stage EM, multi-stage EM, and multi-stage, generalized EM (GEM).

iterations. The likelihood is not guaranteed to increase at each iteration if the fidelity of the orbit model changes, but a low-fidelity orbit model may move the elements towards values with greater likelihood where using a higher-fidelity model may be more beneficial. This is shown in Fig. 16, where data is simulated for four objects and clutter as before, but now J_2 perturbations from Earth's oblateness are modeled [57]. For the first thirty iterations of six different EM initializations, a Keplerian orbit model is assumed and the elements estimated as before. Since the Keplerian orbit model can quickly be evaluated, this first set of EM iterations improves the initial orbit estimates without significant computational cost. The orbit model is then changed to include the perturbations, and the position and velocity of each object at the initial measurement time are estimated; convergence plots for these parameters are given in the supplement. When the orbit model is changed, the concentration parameter is reduced from its current estimate to a new initial value of $\lambda^{(0)} = 10^3$ rad^{-2} to allow the parameters to move through the element space. This creates the sharp decrease in the log-likelihood shown in Fig. 16a. The likelihood eventually recovers after the model changes and runs 3 and 6 reach values close to the likelihood of the actual unknown parameters. Similarly, Fig. 16b shows that the data association accuracy converges to values between 50% and 85% for the Kepler model, but increases for most runs after the orbit model is updated with runs 3 and 6 reaching Rand indices close to 100%. More research can be done on determining how quickly to increase the fidelity of the orbit models and ensuring smoother transitions between models of different fidelities.

The EM algorithms discussed here may not perform well in all tracking scenarios, though these limitations apply to other tracking algorithms as well. When the angular measurement errors are large, the data associations will be innaccurate and the parameter estimates can be biased or have high variance. Alternatively, if the initial orbital-element estimates are poor, then the data association probabilities for the objects will be near-zero and thus the orbital elements will not be updated in the M step. In this case, better initial orbital-element estimates may possibly be obtained with



(a) Log-likelihood

(b) Rand Index

Fig. 16 Results for EM implemented for a Keplerian orbit model for the first 30 iterations and with J_2 perturbations for the next 40 iterations.

the aid of a different object tracking algorithm. EM may also perform poorly if the likelihood function is effectively zero except at the actual orbital elements, which can occur when the concentration parameter is large. This situation will also lead to near-zero data association probabilities that make updating the orbital elements in the M step difficult. To summarize, successful implementation of EM requires good initial starting values for the orbital elements and a likelihood function that allows the M step to move the elements through the parameter space to regions of higher likelihood.

Another limitation of the EM algorithms discussed here are that they only apply to deterministic orbit models. EM algorithms can also be derived for orbit models that include stochastic perturbations that are modeled with process noise, such as atmospheric drag or solar radiation pressure. The resulting statistical model is then a state space model as discussed previously in this section. The basic mixture model framework considered here is not applicable in this case since the likelihood of the measurements is no longer given by (10) due to the temporal dependency induced by the process noise. One approach for adapting EM to state space models is to implement a filter or smoothing algorithm in the E step to obtain the conditional distribution of the object states given the measurements. For a linear, Gaussian model, this is the Kalman filter, and for a non-linear stochastic motion model a particle filter could be used. Parameters in the motion and measurement model are then updated in the M step as for the deterministic orbit model case. Probabilistic MHT (PMHT) is one of the main uses of EM for multiple object tracking when a state space model formulation of the problem is assumed [16, 46]; references for the application of PMHT to space object tracking are given in Section I. However, there are many filtering algorithms for state prediction that account for process noise [58, 74, 75] that have not been incorporated into EM algorithms for space object tracking, and other EM algorithms have been used for multiple object tracking in the state space model framework [34], so constructing new EM algorithms for space object tracking

with these or other filters is potentially of interest as well.

VIII. Conclusion

Expectation-Maximization (EM) is commonly used for probabilistic data association and parameter estimation in a variety of applications. In the space object tracking context, EM associates measurements with objects or clutter and computes maximum likelihood estimates of orbital elements and other mixture model parameters. This paper demonstrated the data association and parameter estimation properties of EM for two different space object tracking scenarios. Simulations showed that estimates of orbital elements and data associations are nearly unbiased at low detection probabilities and that the bias and variance decreases with increasing measurements of the object. The paper also investigated a multi-stage EM approach that is useful when the data includes measurements from a large and possibly unknown number of objects. Recall that multi-stage EM is a procedure to systematically decrease the number of orbit models in the mixture model as orbits are identified, so that the linkage problem becomes simpler and more manageable. In simulations, the procedure accomplished the linkage task more accurately for tens of objects than single-stage EM and correctly identifed up to 50 orbits with high-accuracy and low variance.

Future work on using EM for tracking space objects includes allowing for a larger number of objects, investigating higher-fidelity orbit models and tracking simulators, and replacing the deterministic motion model with a stochastic, non-linear motion model that accounts for perturbations. Furthermore, this paper only considers off-line estimation, but EM algorithms might also be derived for on-line tracking that consist of updating orbital elements as additional data becomes available. For example, when new measurements are obtained, the current orbital-element estimates could be updated by taking them to be starting values for EM and iterating the E and M steps until convergence.

A comparison of EM with more traditional approaches to space object tracking is also needed to establish the method as competitive. Since most alternative methods are filter-based due to their consideration of process noise, a fair comparison will require an EM algorithm that also accounts for process noise. Probabilistic MHT (PMHT) is one EM-based filtering-approach that has been compared to MHT and other tracking methods, but a comparison does not appear to have been made for tracking space objects. Furthermore, filter-based EM algorithms can be derived for tracking that differ from PMHT as demonstrated in [34] or [45], for example. Hence, it will be interesting to compare non-PMHT, filter-based EM algorithms to MHT or PMHT for space object tracking. Note that a filter-based EM algorithm that accounts for process noise will differ from the EM algorithms presented here, but the basic E and M step structure of the algorithm will be the same.

In general, EM algorithms for more complex orbit models and tracking scenarios remain to be derived in detail. The desirable properties and intuitive structure of EM suggests it is likely to be useful for space object tracking scenarios that have not been analyzed yet with EM. We plan to continue investigating the use of EM for these tracking scenarios and models of interest.

Appendix

This appendix describes the mixture model and EM algorithm used to track a single object in clutter in Section VI. The main difference between this EM algorithm and the EM algorithm for tracking multiple objects is that now multiple measurements may be available at each observation time, which changes the detection probability model slightly. The derivation of the two EM algorithms is similar, but the derivation for this second EM algorithm is provided here for completeness.

The data for this scenario are n_i angle measurements taken at time t_i for $i = 1, \ldots, n$. Let $\phi_{ij} = (\phi_{ij,1}, \phi_{ij,2})$ denote the jth angle measurement at the ith time and let Y_{ij} denote the corresponding unit vector for $j = 1, \ldots, n_i$; the measurements at time i are denoted $Y_i = \{Y_{ij}\}_{j=1}^{n_i}$. At each time, at most one measurement is allowed to be of the object. Hence, if no detection occurs at a particular time, then all of the measurements at that time are of clutter. Conversely, if there is a detection, then all but one of the measurements at that time are of clutter. All angle measurements and detections are assumed to be independent for simplicity, though correlation could be modeled here if desired. The parameters to estimate are $\xi = \{p_D, \Theta, \lambda\}$, where p_D is the detection probability and Θ are the orbital elements for the object.

The mixture model for this data is developed as in the multiple object tracking scenario. The latent data association variable is denoted X_i, where $X_i = 0$ indicates no detection was made and $X_i = j$ indicates measurement Y_{ij} is of the object. At each time, a detection occurs with probability p_D, so that the detection probability model is

$$p_\xi(X_i = j) = \begin{cases} 1 - p_D, & j = 0 \\ \\ p_D/n_i, & j = 1, \ldots, n_i. \end{cases} \tag{22}$$

The angle measurement model is

$$p_\xi(y_i | X_i = j) = \begin{cases} p_\xi(y_{ij} | X_i = 0)^{n_i}, & j = 0 \\ \\ p_\xi(y_{ij} | X_i = j) p_\xi(y_{ij} | X_i = 0)^{n_i - 1}, & j = 1, \ldots, n_i, \end{cases} \tag{23}$$

since all n_i measurements are clutter for a non-detection and only the jth measurement is not clutter if there is a detection. For the data in Section VI, the clutter measurement error model is taken to be uniform over a region R with area 0.04 rad^2, so that $p_\xi(y_{ij} | X_i = 0) = 25 I_{\{y_{ij} \in R\}}$. The clutter distribution is assumed known from hypothetical prior knowledge and so does not contain parameters that require estimation. The object measurement error model is taken to be the spherical normal distribution with concentration parameter λ given in (6).

To derive the E step for this EM algorithm, the CDLL for this model is

$$\ell_{X,Y}(\xi) = \log\left(p_\xi(X_{1:n}, Y_{1:n})\right) \tag{24}$$

$$= \sum_{i=1}^{n}\left[\log(p_\xi(Y_i|X_i)) + \log(p_\xi(X_i))\right], \tag{25}$$

where the likelihood terms are given in (23). The E step is to compute the conditional expectation of the CDLL with respect to the measurements,

$$Q(\xi, \xi^{(h-1)}) = E_{\xi^{(h)}}\left[\ell_{X,Y}(\xi)|Y_{1:n}\right] \tag{26}$$

$$= \sum_{i=1}^{n}\sum_{j=0}^{n_i}\left[\log(p_\xi(Y_i|X_i = j)) + \log(p_\xi(X_i = j))\right] w_{ij}^{(h-1)}, \tag{27}$$

where the data association probabilities are derived with Bayes' rule as

$$w_{ij}^{(h-1)} = p_{\xi^{(h-1)}}(X_i = j|Y_{1:n}) \tag{28}$$

$$\propto p_{\xi^{(h-1)}}(Y_i|X_i = j)p_{\xi^{(h-1)}}(X_i = j) \tag{29}$$

and the normalization constant is

$$p_{\xi^{(h-1)}}(Y_i) = \sum_{j=0}^{n_i} p_{\xi^{(h-1)}}(Y_i|X_i = j)p_{\xi^{(h-1)}}(X_i = j). \tag{30}$$

For the M step, the orbital elements are updated by maximizing the function in (26) with respect to Θ; note that the clutter likelihood terms and the mixing probabilities can be ignored since they are independent of Θ. Specifically, the update for the orbital elements is

$$\Theta^{(h)} = \arg\max_{\Theta} \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[-\frac{1}{2}\arccos^2\left(\frac{Y_{ij}^T\tilde{r}(t_i;\Theta)}{\|\tilde{r}(t_i;\Theta)\|}\right)\right] w_{ij}^{(h-1)}. \tag{31}$$

Again, the maximization is a weighted least-squares optimization problem that is solved with the Levenberg-Marquardt algorithm. The update for the probability of detection is

$$p_D^{(h)} = \arg\max_{p_D} \sum_{i=1}^{n}\left[\log(1-p_D)w_{i0}^{(h-1)} + \sum_{j=1}^{n_i}\log\left(\frac{p_D}{n_i}\right) w_{ij}^{(h-1)}\right], \tag{32}$$

which can be accomplished with differentiation and yields the update

$$p_D^{(h)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} w_{ij}^{(h-1)}. \tag{33}$$

The update for the concentration parameter is

$$\lambda^{(h)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} w_{ij}^{(h-1)}}{\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \arccos^2 \left(\frac{Y_{ij}^T \tilde{r}(t_i;\Theta^{(h)})}{\|\tilde{r}(t_i;\Theta^{(h)})\|} \right) w_{ij}^{(h-1)}}, \tag{34}$$

which is derived similarly to the update for λ in (20).

The log-likelihood for a set of mixture model parameters can be evaluated as

$$\ell_Y(\xi) = \sum_{i=1}^{n} \log(p_\xi(Y_i)) \tag{35}$$

$$= \sum_{i=1}^{n} \log \left(\sum_{j=0}^{n_i} p_\xi(Y_i|X_i = j) p_\xi(X_i = j) \right), \tag{36}$$

where the likelihood terms are given in (23) and the prior terms are given in (22). Similar expressions are provided in equations 67-69 in [69].

The analysis in this paper was carried out with the statistical software `R` [76]. In the M step, the Levenberg-Marquardt algorithm provided by the package `minpack.lm` [77] is used for updating the orbital elements. The Rand index is implemented in the package `clusteval` [78] and the imaginary error function needed to evaluate the normalizing constant $C(\lambda)$ is available in the package `pracma` [79], though we approximated this value as discussed in Section III. The likelihood function was written with the package `NIMBLE` [80] and called as compiled C++ code in `R`. Calculating the position can be expensive since it requires solving Kepler's equation to obtain the true anomaly, so doing the computation in C++ significantly reduces the time it takes to evaluate the likelihood function. The package `deSolve` [81] was used to compute the position and velocity for the perturbed orbits considered in Section VII.

Several implementation details of the EM algorithms discussed in this paper should be mentioned since they are not obvious. In the E step, the data association probabilities are scaled to avoid numerical underflow when normalizing their values to sum to one. Consider again the EM algorithm in Section IV, and let $\tilde{w}_{ij}^{(h-1)}$ denote the non-normalized data association probabilities from

$$w_{ij}^{(h-1)} \propto p_{\xi^{(h-1)}}(Y_i|X_i = j)\pi_j^{(h-1)}, \tag{37}$$

where the normalizing constant is $p_{\xi^{(h-1)}}(Y_i) = \sum_{j=0}^{J} \tilde{w}_{ij}^{(h-1)}$. Furthermore, let $\tilde{w}_{i*}^{(h-1)}$ denote the maximum of the $\tilde{w}_{ij}^{(h-1)}$

over j. The logarithm of the normalizing constant can then be computed as

$$\log\left(\sum_{j=0}^{J} \tilde{w}_{ij}^{(h-1)}\right) = \log\left(\tilde{w}_{i*}^{(h-1)}\right) + \log\left(\sum_{j=0}^{J} \exp\left(\log\left(\tilde{w}_{ij}^{(h-1)}\right) - \log\left(\tilde{w}_{i*}^{(h-1)}\right)\right)\right), \tag{38}$$

which is discussed in more detail with relation to the softmax function in [82]. In the M step, using the equinoctial representation of the orbital elements helps avoid parameter identifiability problems, such as when the orbit is circular [59]. Scaling the orbital elements to have similar orders of magnitude may also help avoid numerical problems in the M step.

Funding Sources

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Funding for this work is through the LLNL Laboratory Directed Research and Development (LDRD) Program, (19-SI-004).

Acknowledgments

The author thanks three anonymous reviewers and the editors for comments and suggestions that led to significant improvements in this paper. The author also acknowledges helpful conversations with Robert Armstrong, Will Dawson, Daniel Merl, Joshua Meyers, Caleb Miller, Amanda Muyskens, Eisha Nathan, Benjamin Priest, Geoffrey Sanders, Edward Schlafly, and Michael Schneider.

References

[1] Johnson, N. L., "Space traffic management concepts and practices," *Acta Astronautica*, Vol. 55, No. 3-9, 2004, pp. 803–809. https://doi.org/10.1016/j.actaastro.2004.05.055.

[2] Akella, M. R., and Alfriend, K. T., "Probability of collision between space objects," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 5, 2000, pp. 769–772. https://doi.org/10.2514/2.4611.

[3] Jones, B. A., Bryant, D. S., Vo, B.-T., and Vo, B.-N., "Challenges of multi-target tracking for space situational awareness," *2015 18th International Conference on Information Fusion (Fusion)*, IEEE, 2015, pp. 1278–1285.

[4] Frazer, G., Rutten, M., Cheung, B., and Cervera, M., "Orbit determination using a decametric line-of-sight radar," *Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, 2013.

[5] Davey, S. J., Bessell, T., Cheung, B., and Rutten, M., "Track before detect for space situation awareness," *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2015, pp. 1–7. https://doi.org/10.1109/dicta.2015.7371316.

[6] Xin, W., "Orbit Determination of Mixed Observations of Multiple Objects," *Chinese Astronomy and Astrophysics*, Vol. 39, No. 2, 2015, pp. 254–264. https://doi.org/10.1016/j.chinastron.2015.04.011.

[7] Cheung, B., Rutten, M., Davey, S., and Cohen, G., "Probabilistic multi hypothesis tracker for an event based sensor," *2018 21st International Conference on Information Fusion (FUSION)*, IEEE, 2018, pp. 1–8. https://doi.org/10.23919/ICIF.2018.8455718.

[8] Kent, J. T., Bhattacharjee, S., Hussein, I. I., Faber, W. R., and Jah, M., "Fisher-Bingham-Kent mixture models for angles-only observation processing," *2018 Space Flight Mechanics Meeting*, 2018, p. 1972. https://doi.org/10.2514/6.2018-1972.

[9] Siminski, J., and Jilete, B., "Resolution of Track Association Ambiguities using Expectation-Maximiation Algorithm," *Proc. 1st NEO and Debris Detection Conference*, ESA Space Safety Programme Office, 2019.

[10] Mallick, M., Rubin, S., and Vo, B.-N., "An introduction to force and measurement modeling for space object tracking," *Proceedings of the 16th International Conference on Information Fusion*, IEEE, 2013, pp. 1013–1020.

[11] Bar-Shalom, Y., "Tracking methods in a multitarget environment," *IEEE Transactions on automatic control*, Vol. 23, No. 4, 1978, pp. 618–626. https://doi.org/10.1109/TAC.1978.1101790.

[12] Cox, I. J., "A review of statistical data association techniques for motion correspondence," *International Journal of Computer Vision*, Vol. 10, No. 1, 1993, pp. 53–66. https://doi.org/10.1007/bf01440847.

[13] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with applications to tracking and navigation: theory algorithms and software*, John Wiley & Sons, 2004. https://doi.org/10.1002/0471221279.

[14] Ristic, B., Arulampalam, S., and Gordon, N., "Beyond the Kalman filter," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 19, No. 7, 2004, pp. 37–38.

[15] Pulford, G., "Taxonomy of multiple target tracking methods," *IEE Proceedings-Radar, Sonar and Navigation*, Vol. 152, No. 5, 2005, pp. 291–304. https://doi.org/10.1049/ip-rsn:20045064.

[16] Stone, L. D., Streit, R. L., Corwin, T. L., and Bell, K. L., *Bayesian multiple target tracking*, Artech House, 2013.

[17] Vo, B.-n., Mallick, M., Bar-shalom, Y., Coraluppi, S., Osborne III, R., Mahler, R., and Vo, B.-t., "Multitarget tracking," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1999, pp. 1–15. https://doi.org/10.1002/047134608x.w8275.

[18] Bar-Shalom, Y., Daum, F., and Huang, J., "The probabilistic data association filter," *IEEE Control Systems Magazine*, Vol. 29, No. 6, 2009, pp. 82–100. https://doi.org/10.1109/MCS.2009.934469.

[19] Stauch, J., Jah, M., Baldwin, J., Kelecy, T., and Hill, K. A., "Mutual application of joint probabilistic data association, filtering, and smoothing techniques for robust multiple space object tracking," *AIAA/AAS Astrodynamics Specialist Conference*, 2014, p. 4365. https://doi.org/10.2514/6.2014-4365.

[20] Blanding, W., Willett, P., and Bar-Shalom, Y., "ML-PDA: Advances and a new multitarget approach," *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, No. 1, 2007, p. 260186. https://doi.org/10.1155/2008/260186.

[21] Reid, D., "An algorithm for tracking multiple targets," *IEEE transactions on Automatic Control*, Vol. 24, No. 6, 1979, pp. 843–854. https://doi.org/10.1109/TAC.1979.1102177.

[22] Blackman, S. S., "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 19, No. 1, 2004, pp. 5–18. https://doi.org/10.1109/MAES.2004.1263228.

[23] Singh, N., Horwood, J. T., Aristoff, J. M., Poore, A. B., Sheaff, C., and Jah, M. K., "Multiple hypothesis tracking (MHT) for space surveillance: Results and simulation studies," Tech. rep., Air Force Research Lab Kirtland AFB NM Space Vehicles Directorate, 2013.

[24] Aristoff, J. M., Beach, D. J. C., Ferris, A., Horwood, J. T., Mont, A. D., Singh, N., and Poore, A. B., "Multiple Frame Assignment Space Tracker (MFAST): Results on UCT Processing," *2015 AIAA/AAS Astrodynamics Specialist Conference, Vail, CO*, 2015. (Paper AAS-15-675).

[25] Vo, B.-N., Singh, S., and Doucet, A., "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Transactions on Aerospace and electronic systems*, Vol. 41, No. 4, 2005, pp. 1224–1245. https://doi.org/10.1109/taes.2005.1561884.

[26] DeMars, K. J., Hussein, I. I., Frueh, C., Jah, M. K., and Scott Erwin, R., "Multiple-object space surveillance tracking using finite-set statistics," *Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 9, 2015, pp. 1741–1756. https://doi.org/10.2514/1.g000987.

[27] McCabe, J. S., DeMars, K. J., and Frueh, C., "Integrated detection and tracking for multiple space objects," *Proceedings of the AAS/AIAA 24th Space Flight Mechanics Meeting*, Univelt, Inc. San Diego, CA, 2015.

[28] Vo, B.-N., and Vo, B.-T., "Random finite set multi-target trackers: stochastic geometry for space situational awareness," *Signal Processing, Sensor/Information Fusion, and Target Recognition XXIV*, Vol. 9474, International Society for Optics and Photonics, 2015, p. 94740H. https://doi.org/10.1117/12.2180839.

[29] Wei, B., and Nener, B., "Distributed space debris tracking with consensus labeled random finite set filtering," *Sensors*, Vol. 18, No. 9, 2018, p. 3005. https://doi.org/10.3390/s18093005.

[30] LeGrand, K. A., and DeMars, K. J., "Relative multiple space object tracking using intensity filters," *2015 18th International Conference on Information Fusion (Fusion)*, IEEE, 2015, pp. 1253–1261.

[31] Delande, E., Frueh, C., Franco, J., Houssineau, J., and Clark, D., "Novel multi-object filtering approach for space situational awareness," *Journal of Guidance, Control, and Dynamics*, Vol. 41, No. 1, 2017, pp. 59–73. https://doi.org/10.2514/1.g002067.

[32] Delande, E., Houssineau, J., Franco, J., Frueh, C., Clark, D., and Jah, M., "A new multi-target tracking algorithm for a large number of orbiting objects," *Advances in Space Research*, Vol. 64, No. 3, 2019, pp. 645–667. https://doi.org/10.1016/j.asr.2019.04.012.

[33] Schneider, M. D., "Bayesian linking of geosynchronous orbital debris tracks as seen by the Large Synoptic Survey Telescope," *Advances in Space Research*, Vol. 49, No. 4, 2012, pp. 655–666. https://doi.org/10.1016/j.asr.2011.11.011.

[34] Shumway, R. H., and Stoffer, D. S., "Dynamic linear models with switching," *Journal of the American Statistical Association*, Vol. 86, No. 415, 1991, pp. 763–769. https://doi.org/10.1080/01621459.1991.10475107.

[35] Avitzour, D., "A maximum likelihood approach to data association," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 2, 1992, pp. 560–566. https://doi.org/10.1109/7.144581.

[36] Gauvrit, H., Le Cadre, J.-P., and Jauffret, C., "A formulation of multitarget tracking as an incomplete data problem," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 33, No. 4, 1997, pp. 1242–1257. https://doi.org/10.1109/7.625121.

[37] Molnar, K. J., and Modestino, J. W., "Application of the EM algorithm for the multitarget/multisensor tracking problem," *IEEE Transactions on Signal Processing*, Vol. 46, No. 1, 1998, pp. 115–129. https://doi.org/10.1109/78.651193.

[38] Bergman, N., "Recursive Bayesian estimation: Navigation and tracking applications," Ph.D. thesis, Linköping University, 1999.

[39] Frenkel, L., and Feder, M., "Recursive expectation-maximization (EM) algorithms for time-varying parameters with applications to multiple target tracking," *IEEE Transactions on Signal Processing*, Vol. 47, No. 2, 1999, pp. 306–320. https://doi.org/10.1109/78.740104.

[40] Malyutov, M., Lu, M., Nikiforov, A., and Protassov, R., "Multitrajectory estimation in noise and clutter," *Proc. 4th Conf. Info. Fusion, Montréal*, 2001, pp. 17–25.

[41] Logothetis, A., Krishnamurthy, V., and Holst, J., "A Bayesian EM algorithm for optimal tracking of a maneuvering target in clutter," *Signal processing*, Vol. 82, No. 3, 2002, pp. 473–490. https://doi.org/10.1016/s0165-1684(01)00198-0.

[42] Davey, S., Colegrove, S., and Gray, D., "A comparison of track initiation with PDAF and PMHT," *2002 International Radar Conference (Radar 2002)*, IET, 2002. https://doi.org/10.1049/cp:20020300.

[43] Davey, S. J., *Extensions to the probabilistic multi-hypothesis tracker for improved data association*, Citeseer, 2003.

[44] Crouse, D. F., Guerriero, M., and Willett, P., "A Critical Look at the PMHT," *J. Adv. Inf. Fusion*, Vol. 4, No. 2, 2009, pp. 93–116.

[45] Yıldırım, S., Jiang, L., Singh, S. S., and Dean, T. A., "Calibrating the Gaussian multi-target tracking model," *Statistics and Computing*, Vol. 25, No. 3, 2015, pp. 595–608. https://doi.org/10.1007/s11222-014-9456-2.

[46] Streit, R. L., and Luginbuhl, T. E., "Probabilistic multi-hypothesis tracking," Tech. rep., Naval Underwater Systems Center, Newport, RI, 1995.

[47] Lan, H., Wang, X., Pan, Q., Yang, F., Wang, Z., and Liang, Y., "A survey on joint tracking using expectation–maximization based techniques," *Information Fusion*, Vol. 30, 2016, pp. 52–68. https://doi.org/10.1016/j.inffus.2015.11.008.

[48] Davey, S. J., and Gaetjens, H. X., *Track-Before-Detect Using Expectation Maximisation: The Histogram Probabilistic Multi-hypothesis Tracker: Theory and Applications*, Springer, 2018. https://doi.org/10.1007/978-981-10-7593-3.

[49] Schoenecker, S., Willett, P., and Bar-Shalom, Y., "Maximum likelihood probabilistic multi-hypothesis tracker applied to multistatic sonar data sets," *Signal Processing, Sensor Fusion, and Target Recognition XX*, Vol. 8050, International Society for Optics and Photonics, 2011, p. 80500A. https://doi.org/10.1117/12.884766.

[50] McCabe, J. S., and DeMars, K. J., "Particle filter methods for space object tracking," *AIAA/AAS Astrodynamics Specialist Conference*, 2014, p. 4308. https://doi.org/10.2514/6.2014-4308.

[51] Wei, B., and Nener, B. D., "Multi-sensor space debris tracking for space situational awareness with labeled random finite sets," *IEEE Access*, Vol. 7, 2019, pp. 36991–37003. https://doi.org/10.1109/access.2019.2904545.

[52] Hauberg, S., "Directional Statistics with the Spherical Normal Distribution," *2018 21st International Conference on Information Fusion (FUSION)*, IEEE, 2018, pp. 704–711. https://doi.org/10.23919/icif.2018.8455242.

[53] Coraluppi, S., and Carthel, C., "Multi-Stage Multiple-Hypothesis Tracking," *J. Adv. Inf. Fusion*, Vol. 6, No. 1, 2011, pp. 57–67.

[54] Rand, W. M., "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, Vol. 66, No. 336, 1971, pp. 846–850. https://doi.org/10.1080/01621459.1971.10482356.

[55] Redner, R. A., and Walker, H. F., "Mixture densities, maximum likelihood and the EM algorithm," *SIAM review*, Vol. 26, No. 2, 1984, pp. 195–239. https://doi.org/10.1137/1026034.

[56] Ng, S. K., Krishnan, T., and McLachlan, G. J., "The EM algorithm," *Handbook of computational statistics*, Springer, 2012, pp. 139–172. https://doi.org/10.1007/978-3-642-21551-3.

[57] Vallado, D. A., *Fundamentals of astrodynamics and applications*, Vol. 12, Springer Science & Business Media, 2001.

[58] Teixeira, B. O., Santillo, M. A., Erwin, R. S., and Bernstein, D. S., "Spacecraft tracking using sampled-data Kalman filters," *IEEE Control Systems Magazine*, Vol. 28, No. 4, 2008, pp. 78–94. https://doi.org/10.1109/mcs.2008.923231.

[59] Broucke, R. A., and Cefola, P. J., "On the equinoctial orbit elements," *Celestial mechanics*, Vol. 5, No. 3, 1972, pp. 303–310. https://doi.org/10.1007/BF01228432.

[60] Kent, J. T., Hussein, I., and Jah, M. K., "Directional distributions in tracking of space debris," *2016 19th International Conference on Information Fusion (FUSION)*, IEEE, 2016, pp. 2081–2086.

[61] Storlie, C. B., Lee, T. C., Hannig, J., and Nychka, D., "Tracking of multiple merging and splitting targets: A statistical perspective," *Statistica Sinica*, 2009, pp. 1–31.

[62] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39, No. 1, 1977, pp. 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

[63] McLachlan, G., and Krishnan, T., *The EM algorithm and extensions*, Vol. 382, John Wiley & Sons, 2007. https://doi.org/10.1002/9780470191613.

[64] Dhillon, I. S., and Sra, S., "Modeling data using directional distributions," Tech. rep., Technical Report TR-03-06, Department of Computer Sciences, The University of Texas at Austin, 2003.

[65] Hornik, K., and Grün, B., "movMF: an R package for fitting mixtures of von Mises-Fisher distributions," *Journal of Statistical Software*, Vol. 58, No. 10, 2014, pp. 1–31. https://doi.org/10.18637/jss.v058.i10.

[66] Wu, C. J., "On the convergence properties of the EM algorithm," *The Annals of statistics*, Vol. 11, No. 1, 1983, pp. 95–103. https://doi.org/10.1214/aos/1176346060.

[67] Moré, J. J., "The Levenberg-Marquardt algorithm: implementation and theory," *Numerical analysis*, Springer, 1978, pp. 105–116. https://doi.org/10.1007/bfb0067700.

[68] Bishop, C. M., *Pattern recognition and machine learning*, Springer, 2006.

[69] Kirubarajan, T., and Bar-Shalom, Y., "Probabilistic data association techniques for target tracking in clutter," *Proceedings of the IEEE*, Vol. 92, No. 3, 2004, pp. 536–557. https://doi.org/10.1109/jproc.2003.823149.

[70] Li, X. R., and Bar-Shalom, Y., "Tracking in clutter with nearest neighbor filters: analysis and performance," *IEEE transactions on aerospace and electronic systems*, Vol. 32, No. 3, 1996, pp. 995–1010. https://doi.org/10.1109/7.532259.

[71] Pulford, G. W., and Logothetis, A., "An expectation-maximisation tracker for multiple observations of a single target in clutter," *Proceedings of the 36th IEEE Conference on Decision and Control*, Vol. 5, IEEE, 1997, pp. 4997–5003. https://doi.org/10.1109/cdc.1997.649846.

[72] Liu, W., Wen, C., Han, C., and Lian, F., "A Bayesian estimation for single target tracking based on state mixture models," *Signal Processing*, Vol. 92, No. 7, 2012, pp. 1706–1714. https://doi.org/10.1016/j.sigpro.2012.01.006.

[73] Kimura, T., Tokuda, T., Nakada, Y., Nokajima, T., Matsumoto, T., and Doucet, A., "Expectation-maximization algorithms for inference in Dirichlet processes mixture," *Pattern Analysis and Applications*, Vol. 16, No. 1, 2013, pp. 55–67. https://doi.org/10.1007/s10044-011-0256-4.

[74] Jazwinski, A. H., *Stochastic processes and filtering theory*, Courier Corporation, 2007.

[75] Hartikainen, J., Seppanen, M., and Sarkka, S., "State-space inference for non-linear latent force models with application to satellite orbit prediction," *arXiv preprint arXiv:1206.4670*, 2012.

[76] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org.

[77] Elzhov, T. V., Mullen, K. M., Spiess, A., and Bolker, B., "R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, Plus Support for Bounds," 2010.

[78] Ramey, J. A., *clusteval: Evaluation of Clustering Algorithms*, 2012. URL https://CRAN.R-project.org/package=clusteval, r package version 0.1.

[79] Borchers, H. W., *pracma: Practical Numerical Math Functions*, 2019. URL https://CRAN.R-project.org/package=pracma, r package version 2.2.9.

[80] de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., and Bodik, R., "Programming with models: writing statistical algorithms for general model structures with NIMBLE," *Journal of Computational and Graphical Statistics*, Vol. 26, 2017, pp. 403–413. https://doi.org/10.1080/10618600.2016.1172487.

[81] Soetaert, K. E., Petzoldt, T., and Setzer, R. W., "Solving differential equations in R: package deSolve," *Journal of statistical software*, Vol. 33, 2010. https://doi.org/10.18637/jss.v033.i09.

[82] Goodfellow, I., Bengio, Y., and Courville, A., *Deep learning*, MIT press, 2016.