Gaussian Mixture Nonlinear Filtering With Resampling for Mixand Narrowing

Mark L. Psiaki

***Abstract*—A new Gaussian mixture filter has been developed, one that uses a resampling step in order to limit the covariances of its individual Gaussian components. The new filter has been designed to produce accurate solutions of difficult nonlinear/non-Gaussian estimation problems. It uses static multiple-model filter calculations and extended Kalman filter (EKF) approximations for each Gaussian mixand in order to perform dynamic propagation and measurement update. The resampling step uses a newly designed algorithm that employs matrix inequalities in order to bound each mixand's covariance. Resampling occurs between the dynamic propagation and the measurement update in order to ensure bounded covariance in both of these operations. The resampling algorithm also attempts to economize on the number of new mixands. The resulting filter has been tested on a difficult seven-state nonlinear filtering problem. It achieves significantly better accuracy than a simple EKF, an unscented Kalman filter, a moving-horizon estimator/backwards-smoothing EKF, and a regularized particle filter.**

***Index Terms*—Bayesian filter, Gaussian mixture filter, Kalman filter, nonlinear filter.**

I. Introduction

KALMAN filters have been applied to Bayesian estimation problems with stochastic dynamic models and noisy measurements for over 5 decades [1], [2]. The original formulation solved the discrete-time problem with linear dynamics, linear measurements, and Gaussian noise and priors [1]. The general discrete-time nonlinear problem, however, has no known closed-form solution. A number of approximate solutions for nonlinear/non-Gaussian problems have been developed, and they have met with varying degrees of success. These include the extended Kalman filter (EKF) [3]–[5]—perhaps better characterized as the family of EKFs [6], the unscented or sigma-points Kalman filter (UKF) [7], [8], the family of filters known as particle filters (PFs) [5], [9], the moving horizon estimator [10]—aka the backward-smoothing extended Kalman filter (BSEKF) [11], and a family known as Gaussian mixture filters [12]–[16].

There are two general ways in which these approximate filters can perform poorly. They can diverge, or they may yield suboptimal or in-consistent estimation accuracy. These two failure modes are illustrated in Ref. [17] via simulation tests that use a benchmark 7-state nonlinear estimation problem, the blind tricyclist problem. The UKF can diverge, and its accuracy is very poor. The EKF and PF sometimes nearly diverge, and their accuracies are far from the Cramer-Rao lower bound. The BSEKF converges and has the best estimation accuracy, but its accuracy is not very close to the Cramer-Rao lower bound. Furthermore, the BSEKF is expensive in terms of needed computational resources, and the PF is even more expensive.

Reference [17] did not attempt to include a Gaussian mixture filter in its performance comparison due to a known problem of such filters: If their component mixands have covariances that are too large, then the EKF or UKF component calculations will not have sufficient accuracy to achieve good performance [14], [15], [18], [19].

This paper's main contribution is to develop a new type of Gaussian mixture filter that avoids the problem of having mixand covariances which are too large. Its algorithm uses multiple-model filter calculations that are based on first-order EKF approximations carried out on a mixand-by-mixand basis. The measurement update involves re-weighting of the mixands using the usual innovations-based technique of static multiple-model filters [4]. In addition to these standard Gaussian mixture filter operations, a mixture re-sampling/re-approximation step is added between the dynamic propagation and the measurement update. This re-sampling step uses the new Gaussian mixture re-approximation algorithm of Ref. [20]. This latter algorithm starts with an input Gaussian mixture and a matrix inequality (MI) upper bound that will apply to the covariance matrices of the output mixands. If all of the original mixands respect this bound, then the output mixture essentially equals the input mixture, except that importance-type re-sampling properties will act to eliminate mixands with very low weights. If any of the original mixands do not respect the MI covariance upper bound, then the re-sampling algorithm constructs a new Gaussian mixture that approximates the original mixture accurately while enforcing the MI bound. The user selects an upper bound which is small enough to ensure that first-order EKF measurement update and dynamic propagation operations can be applied to each mixand with very little truncation error. This constraint ensures the overall accuracy with which the Gaussian mixture filter approximates the exact Bayesian posterior distribution.

The present paper is closely coupled to Ref. [20]. Its contribution is to incorporate Ref. [20]'s new mixture re-approximation algorithm into a full nonlinear filter. Without the present paper, it would be unclear how the new Gaussian mixture re-sampling algorithm of Ref. [20] should be used to develop a full nonlinear filter.

This nonlinear filtering approach is akin to a PF, but with two important distinctions. First, its re-sampling algorithm approximates one Gaussian mixture by constructing another, as opposed to approximating one ensemble of particles by

Manuscript received September 4, 2015; revised May 5, 2016 and June 20, 2016; accepted July 2, 2016. Date of publication July 27, 2016; date of current version September 8, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tongtong Li.

The author is with the Deptartment of Aerospace and Ocean Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: mlpsiaki@vt.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2595503

constructing another. Second, the re-sampling operation occurs in a different place than for a typical PF, before the measurement update rather than after it [5], [9]. This alternate position enables a single re-sampling step to enforce the filter's MI covariance bound for both the measurement update of the current sample and the dynamic propagation of the next sample.

In all other respects, this filter is very similar to a PF, except that the PF's infinitesimal-width particles are replaced by Gaussian mixands that have non-infinitesimal widths. Therefore, one might call each mixand a "blob" in contrast to the term "particle" that is used for an ensemble element of a PF. This contrast suggests the term "Blob Filter" as a descriptive means of naming Gaussian mixture filters that employ such a mixture re-approximation step.

This new Gaussian mixture filter differs from existing Gaussian mixture filters in several respects. The filters described in Refs. [12] and [16] include neither a re-sampling step nor any other means to limit mixand covariances. The filter in Ref. [13] includes re-sampling, but not for the purpose of enforcing a covariance upper limit on each mixand. Instead, re-sampling is applied primarily to limit the number of mixands, which is somewhat analogous to the action of importance re-sampling in a PF to prune away particles with low weight. References [14], [15], [18], and [19] develop filters or parts of filters that incorporate Gaussian mixture re-approximation with the goal of limiting the mixand covariances. All of these methods rely on re-approximation of a given wide 1-dimensional Gaussian by a pre-computed mixture of narrower weighted 1-dimensional Gaussians. Re-approximation of a multi-dimensional Gaussian mixture is performed using products of mixture re-approximations of 1-dimensional Gaussians along principal axes of each original mixand's covariance matrix. This technique has two drawbacks. First, the number of needed new mixands in the re-approximation is exponential in the problem dimension. Second, there is no obvious way to exploit the overlap of original wide mixands in order to conserve on the number of narrowed mixands needed in the re-approximated distribution.

Although Gaussian mixture filters have been around for decades, Ref. [12], the idea of limiting their mixand covariances to achieve accurate nonlinear filtering has been developed only recently. The relative newness of this technique has left open questions about how best to incorporate covariance bounds, [14], [15], [18], [19]. Reference [14] uses two mixtures, a nominal mixture and a second that has smaller mixand covariances and, presumably, more mixands. If they are not close enough in a functional norm sense, then refinement takes place. The mixture with smaller covariances becomes the nominal, and a new fine-scale mixture is generated by further splitting of mixands. If the two mixtures are too close, then the nominal becomes the new fine-scale mixture, and mixands of the original nominal are merged in order to generate the new nominal mixture. Nothing is said about where this mixture adaptation occurs in the sequence of calculations for a given sample interval. No mention is made of the fact that some mixands may need coarsening while others may need refinement. This method may be expensive computationally because of its need to propagate and update two mixtures simultaneously.

Reference [15] is targeted at the application of satellite orbit determination. It does not propose a general Gaussian mixture re-approximation scheme. Rather, it limits mixand variances only in the direction of orbital semi-major axis and only once, at the initial time. The one significant adaptive element of its filter is a weight adaptation, similar to Ref. [16], which is a completely different technique than that used in the present paper and typically a less effective technique.

Reference [18] deals with Gaussian mixture re-approximation only for nonlinear dynamic propagation. It says nothing about refinement for purposes of dealing with measurement nonlinearities. A practitioner developing a filter for a system with a linear dynamic model and nonlinear measurements would not be helped by Ref. [18]. He or she would not learn how or when to use a Gaussian mixture re-approximation in order to perform Bayesian measurement updates accurately in the presence of strong measurement nonlinearities.

Reference [19] concentrates on the splitting of mixands that are too wide. It says little about the merging of mixands that nearly overlap. It never explicitly defines a nonlinear filter that uses its mixand splitting technique. Instead, it gives an example filter calculation that includes just one dynamic propagation step followed by one measurement update. It performs the splitting—its version of mixture re-approximation—prior to the dynamic propagation, which is not the location of the present paper's re-approximation step. Also, Ref. [19] envisions the merging of mixands and the splitting of mixands as two distinct operations that may occur at different points of the filtering process, which is different from the strategy presented in the current paper.

Thus, the current paper defines a new Gaussian mixture nonlinear filtering algorithm that employs mixture re-approximation. Its algorithm is not found in or derivable from Refs. [14], [15], [18], or [19].

Another contribution of this paper is a new square-root information filter (SRIF) formulation of the static multiple-model filter calculations. These calculations form the basis of the new Gaussian mixture filter. Versions of the needed calculations have been published in multiple places for covariance filter implementations, e.g., Ref. [4], but this paper gives the first known SRIF formulation. The Gaussian mixture re-approximation algorithm of Ref. [20] is also developed in the square-root information form. Therefore, the present paper's development of the SRIF multiple-model filter is needed in order to make the mixture re-approximation algorithm of Ref. [20] directly usable within a Gaussian mixture filtering framework.

An additional contribution of the present paper is a simulation-based evaluation of the new Gaussian mixture filter on the blind tricyclist nonlinear estimation problem [17]. Other Gaussian mixture papers evaluate their proposed filters on problems of much lower dimension, typically 2 or 3, or they evaluate only a single prediction step. The present work tests the new filter on the 7-state blind tricyclist problem and compares its performance and computational cost to those of an EKF, two UKFs, two BSEKFs, and two PFs. The new filter's accuracy is also compared to the problem's Cramer-Rao lower bound. The author has successfully applied the new filter to problems of practical importance too, including the 12-state spacecraft attitude/attitude-rate/moment-of-inertia estimation problem of

[11] and a 6-state spacecraft orbit determination problem. The blind tricyclist problem is used here because of the documented comparisons with other filters available in [17].

This paper develops and evaluates its new Gaussian mixture Bayesian filter in five main sections. Section II poses the discrete-time nonlinear/non-Gaussian dynamic filtering problem that will be solved using a Gaussian mixture filter. It also reviews the problem's theoretical Bayesian solution. Section III defines various Gaussian mixtures using square-root information matrix notation. Section IV defines a MI that bounds the mixand covariances of a re-sampled Gaussian mixture, and it reviews the associated new algorithm for mixture re-sampling that enforces this bound. Section V develops the full Gaussian mixture filtering algorithm that uses the new re-sampling technique. Section VI performs Monte-Carlo simulation tests of the new algorithm's performance on the blind tricyclist problem, and it compares this performance with that of other filters. Section VII gives this paper's conclusions.

II. Discrete-Time Nonlinear Filtering Problem Definition and Theoretical Solution

A. Filtering Problem Definition

The discrete-time nonlinear/non-Gaussian Bayesian filtering problem includes a dynamics model and a measurement model. They take the respective forms:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k), && \text{for } k = 0, 1, 2, \ldots \\ \boldsymbol{y}_{k+1} &= \boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1}) + \boldsymbol{\nu}_{k+1}, && \text{for } k = 0, 1, 2, \ldots \end{aligned} \tag{1}$$

where k is the discrete-time sample index, $\boldsymbol{x}_k$ is the n_x-dimensional state vector at sample k, $\boldsymbol{w}_k$ is the n_w-dimensional process noise vector that applies for the state transition from sample k to sample $k+1$, $\boldsymbol{y}_{k+1}$ is the n_y-dimensional measurement vector at sample $k+1$, and $\boldsymbol{\nu}_{k+1}$ is the n_y-dimensional measurement noise vector at sample $k+1$. The vector function $\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$ is the nonlinear discrete-time dynamics state transition function, and the vector function $\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})$ is the nonlinear measurement function. These two functions are assumed to be continuous with continuous first derivatives.

The definition of 3 *a priori* probability density functions completes the filtering problem formulation. The first is the *a priori* probability density for $\boldsymbol{x}_0$. Let it be defined as $p_{x_0}(\boldsymbol{x}_0)$. The second is the *a priori* probability density function for each $\boldsymbol{w}_k$. Its definition is $p_{w_k}(\boldsymbol{w}_k)$. The third is the *a priori* probability density function for each $\boldsymbol{\nu}_{k+1}$, which is defined as $p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1})$. Let the probability density functions $p_{w_k}(\boldsymbol{w}_k)$ and $p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1})$ be defined for all $k = 0, 1, 2, \ldots$.

In principle, the new filter can work with any sufficiently smooth nonlinear model functions $\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$ and $\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})$. In practice, the efficiency of the new filter will be impacted by the severity of the nonlinearities. If the nonlinearities are strong over regions of high probability density, then the underlying Gaussian mixture approximations will need to be split into many mixands with very small individual covariances. If the dimension of the state space is high with a high-dimensional subspace that has strong nonlinearities, then the number of required mixands can become too large for practical operation. The question of when a problem becomes numerically intractable is difficult to answer without first trying the method on the given problem.

B. Theoretical Bayesian Solution

The theoretical solution to the filtering problem is the conditional probability density function $p_{x_k}(\boldsymbol{x}_k|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)$. This is the probability density of $\boldsymbol{x}_k$ conditioned on all the data starting from the initial measurement $\boldsymbol{y}_1$ and extending up to the measurement $\boldsymbol{y}_k$, which is the last one available at the applicable sample time of $\boldsymbol{x}_k$. This is commonly known as the *a posteriori* probability density function. Let this function be written in short-hand form as $p_{x_k}(\boldsymbol{x}_k|k)$, and let the definition of this latter function be extended to include $p_{x_0}(\boldsymbol{x}_0|0) = p_{x_0}(\boldsymbol{x}_0)$.

An additional useful probability density function is the *a priori* state probability density, $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)$. Let this latter function be designated by the short-hand form $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$.

The Bayesian filtering operations start at sample $k = 0$ and successively compute $p_{x_1}(\boldsymbol{x}_1|0)$, $p_{x_1}(\boldsymbol{x}_1|1)$, $p_{x_2}(\boldsymbol{x}_2|1)$, $p_{x_2}(\boldsymbol{x}_2|2)$, $p_{x_3}(\boldsymbol{x}_3|2)$, $p_{x_3}(\boldsymbol{x}_3|3)$, ... These computations can be completely characterized by a recursion that starts with $p_{x_k}(\boldsymbol{x}_k|k)$, $p_{w_k}(\boldsymbol{w}_k)$, and $p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1})$ and that uses the dynamics and measurement models in Eq. (1) to compute $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ followed by $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$. These computations start by forming the following probability density functions:

$$\begin{aligned} &p_{x_{k+1}}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{w}_k) \\ &\quad = \delta\left[\boldsymbol{x}_{k+1} - \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)\right] \end{aligned} \tag{2}$$

$$\begin{aligned} &p_{x_{k+1}x_k w_k}(\boldsymbol{x}_{k+1}, \boldsymbol{x}_k, \boldsymbol{w}_k|k) \\ &\quad = p_{x_{k+1}}(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{w}_k)\, p_{x_k w_k}(\boldsymbol{x}_k, \boldsymbol{w}_k|k) \\ &\quad = \delta\left[\boldsymbol{x}_{k+1} - \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)\right] p_{x_k}(\boldsymbol{x}_k|k)\, p_{w_k}(\boldsymbol{w}_k) \end{aligned} \tag{3}$$

where $\delta[]$ is a Dirac delta function that takes a vector argument which is a perturbation of the $\boldsymbol{x}_{k+1}$ vector away from the output of the dynamics model function $\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$.

The $\boldsymbol{x}_k$ and $\boldsymbol{w}_k$ dependence of the probability density function in Eq. (3) is integrated out to yield the desired *a priori* probability density function at sample $k+1$:

$$\begin{aligned} &p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k) \\ &= \int_{-\infty}^{\infty} d\boldsymbol{w}_k \int_{-\infty}^{\infty} d\boldsymbol{x}_k p_{x_{k+1}x_k w_k}(\boldsymbol{x}_{k+1}, \boldsymbol{x}_k, \boldsymbol{w}_k|k) \\ &= \int_{-\infty}^{\infty} d\boldsymbol{w}_k p_{w_k}(\boldsymbol{w}_k) \int_{-\infty}^{\infty} d\boldsymbol{x}_k \delta\left[\boldsymbol{x}_{k+1} - \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)\right] p_{x_k}(\boldsymbol{x}_k|k) \\ &= \int_{-\infty}^{\infty} d\boldsymbol{w}_k p_{w_k}(\boldsymbol{w}_k) \frac{p_{x_k}\left[\boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1}, \boldsymbol{w}_k)|k\right]}{|\det(\Phi_k)|} \end{aligned} \tag{4}$$

where the function $\boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1}, \boldsymbol{w}_k)$ in the last line of Eq. (4) is the inverse of the dynamics function in the first line of Eq. (1). It is defined so that $\boldsymbol{x}_{k+1} = \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$ and $\boldsymbol{x}_k = \boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1}, \boldsymbol{w}_k)$ are equivalent, which is the same as saying that the latter equation is valid for all combinations of

$\boldsymbol{x}_{k+1}$, $\boldsymbol{x}_k$, and $\boldsymbol{w}_k$ that obey the former equation. The scalar function det() returns the determinant of its matrix input argument. The n_x-by-n_x matrix Φ_k in Eq. (4) is the Jacobian first partial derivative of the dynamics function:

$$\Phi_k = \left.\frac{\partial \boldsymbol{f}_k}{\partial \boldsymbol{x}_k}\right|_{\left[\boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1},\boldsymbol{w}_k),\boldsymbol{w}_k\right]}. \tag{5}$$

The final operations of the Bayesian recursion require the measurement probability density conditioned on the state at sample $k+1$:

$$p_{y_{k+1}}\left(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}\right) = p_{\nu_{k+1}}\left[\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})\right]. \tag{6}$$

The measurement update is then

$$\begin{aligned} &p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1) \\ &= \frac{p_{y_{k+1}}(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1})p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)}{\int_{-\infty}^{\infty} d\check{\boldsymbol{x}}_{k+1} p_{y_{k+1}}(\boldsymbol{y}_{k+1}|\check{\boldsymbol{x}}_{k+1})p_{x_{k+1}}(\check{\boldsymbol{x}}_{k+1}|k)} \\ &= \frac{p_{\nu_{k+1}}\left[\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})\right] p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)}{\int_{-\infty}^{\infty} d\check{\boldsymbol{x}}_{k+1} p_{\nu_{k+1}}\left[\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\check{\boldsymbol{x}}_{k+1})\right] p_{x_{k+1}}(\check{\boldsymbol{x}}_{k+1}|k)}. \end{aligned} \tag{7}$$

Unfortunately, many of the theoretical calculations given in the dynamic propagation in Eq. (4) and given in the measurement update in Eq. (7) cannot be evaluated in closed form. The inverse dynamics function $\boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1},\boldsymbol{w}_k)$ often cannot be derived analytically. The last line of Eq. (4) is unlikely to allow closed-form evaluation of its integral with respect to $\boldsymbol{w}_k$. The integral in the denominator of the second line of Eq. (7) would likely not allow closed-form computation. Even if one were able to produce closed-form results for these steps, the resulting $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$ likely would be too complicated to allow closed-form computation of the *a posteriori* expected value of $\boldsymbol{x}_{k+1}$ or its covariance. Furthermore, the resulting $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$ would constitute an input to the Eq. (4) $\boldsymbol{w}_k$ integral for the next sample interval. In all but some special cases, one eventually loses the ability to derive closed-form expressions for the probability density functions of interest.

These difficulties necessitate the development of approximate solutions to this nonlinear/non-Gaussian Bayesian estimation problem. The present approach uses Gaussian mixtures to approximate all of these probability density functions. Given mixtures with individual elements that have sufficiently small covariances, local linearized approximations of the functions $\boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1},\boldsymbol{w}_k)$ and $\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})$ can be used. All of the needed integrals are computable in closed form when working with Gaussian mixture probability density functions and local linearizations of the problem model functions.

III. Definition of Gaussian Mixture Probability Density Functions

A. Gaussian Mixture in Square-Root Information Form

A Gaussian mixture probability density function is a weighted sum of individual Gaussian terms. A Gaussian mixture can be written in the following square-root information form

$$p(\boldsymbol{x}) = \sum_{i=1}^{N} \mathtt{w}_i \mathcal{N}_{sr}(\boldsymbol{x};\boldsymbol{\mu}_i, R_i) \tag{8}$$

where the scalars $\mathtt{w}_i$ for $i = 1, \ldots, N$ are the mixand weights and the functions

$$\begin{aligned} \mathcal{N}_{sr}(\boldsymbol{x};\boldsymbol{\mu}_i, R_i) = \frac{|\det(R_i)|}{(2\pi)^{n_x/2}} e^{-0.5[R_i(\boldsymbol{x}-\boldsymbol{\mu}_i)]^{\mathrm{T}}[R_i(\boldsymbol{x}-\boldsymbol{\mu}_i)]} \\ \text{for } i = 1, \ldots, N \end{aligned} \tag{9}$$

are individual Gaussian distributions for the n_x-dimensional vector $\boldsymbol{x}$. The ith Gaussian mixand is defined by its n_x-dimensional mean value vector $\boldsymbol{\mu}_i$ along with its n_x-by-n_x square-root information matrix R_i. The corresponding n_x-by-n_x covariance matrix of the ith mixand is $P_i = R_i^{-1}R_i^{-\mathrm{T}}$, with $()^{-\mathrm{T}}$ indicating the transpose of the inverse of the given matrix.

The scalar weights of the distribution must be non-negative, and they must sum to 1. Thus, they obey the conditions:

$$\mathtt{w}_i \geq 0, \quad \text{for} \quad i = 1, \ldots, N \quad \text{and } 1 = \sum_{i=1}^{N} \mathtt{w}_i. \tag{10}$$

Gaussian mixtures inherit an important property from their Gaussian components: It is straight-forward to calculate various useful integrals that involve Gaussian mixture probability density functions. For example, it is easy to prove unit-normalization of the integral of a Gaussian mixture by using the unit normalization of each individual Gaussian mixand function $\mathcal{N}_{sr}(\boldsymbol{x};\boldsymbol{\mu}_i, R_i)$ along with the unit normalization summation constraint on the weights $\mathtt{w}_i$, which is given in Eq. (10). Explicit calculations of the mean and covariance of a Gaussian mixture are carried out using similar techniques. The results are:

$$\begin{aligned} \boldsymbol{\mu}_{gm} &= \sum_{i=1}^{N} \mathtt{w}_i \boldsymbol{\mu}_i \\ P_{gm} &= \sum_{i=1}^{N} \mathtt{w}_i \left[R_i^{-1}R_i^{-\mathrm{T}} + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{gm})(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{gm})^{\mathrm{T}}\right]. \end{aligned} \tag{11}$$

B. Specific Gaussian Mixtures Used to Develop the New Filter

Five different Gaussian mixture probability density functions are needed in order to develop this paper's Gaussian mixture filter. One approximates the *a posteriori* state probability density function $p_{x_k}(\boldsymbol{x}_k|k)$. Two are needed to approximate the *a priori* state probability density function $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$, one immediately after the dynamic propagation calculations and a second one after the Gaussian mixture re-sampling operation that bounds mixand covariances. A fourth Gaussian mixture is needed to approximate the *a priori* process noise probability density function $p_{w_k}(\boldsymbol{w}_k)$, and the fifth approximates the *a priori* measurement noise probability density $p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1})$.

Let these approximations be defined as:

$$p_{x_k}(\boldsymbol{x}_k|k) = \sum_{i=1}^{\hat{N}_{x_k}} \hat{\mathtt{w}}_{x_k i} \mathcal{N}_{sr}\left(\boldsymbol{x}_k; \hat{\boldsymbol{\mu}}_{x_k i}, \hat{R}_{x x_k i}\right) \tag{12}$$

$$p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k) = \sum_{i=1}^{\bar{N}_{x_{k+1}}} \bar{\mathtt{w}}_{x_{k+1} i} \mathcal{N}_{sr}\left[\boldsymbol{x}_{k+1}; \bar{\boldsymbol{\mu}}_{x_{k+1} i}, \bar{R}_{x x_{k+1} i}\right] \tag{13}$$

$$\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k) = \sum_{i=1}^{\tilde{N}_{x_{k+1}}} \tilde{\mathtt{w}}_{x_{k+1} i} \mathcal{N}_{sr}\left[\boldsymbol{x}_{k+1}; \tilde{\boldsymbol{\mu}}_{x_{k+1} i}, \tilde{R}_{x x_{k+1} i}\right] \tag{14}$$

$$p_{w_k}(\boldsymbol{w}_k) = \sum_{i=1}^{N_{w_k}} \mathtt{w}_{w_k i} \mathcal{N}_{sr}\left(\boldsymbol{w}_k; \boldsymbol{\mu}_{w_k i}, R_{w w_k i}\right) \tag{15}$$

$$p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1}) = \sum_{i=1}^{N_{\nu_{k+1}}} \mathtt{w}_{\nu_{k+1} i} \mathcal{N}_{sr}\left[\boldsymbol{\nu}_{k+1}; \boldsymbol{\mu}_{\nu_{k+1} i}, R_{\nu \nu_{k+1} i}\right] \tag{16}$$

where each of these five Gaussian mixtures is characterized by its number of mixands and by its sets of mixand weights, mean values, and square-root information matrices. The five Gaussian mixtures' respective mixand counts are $\hat{N}_{x_k}$, $\bar{N}_{x_{k+1}}$, $\tilde{N}_{x_{k+1}}$, N_{w_k}, and $N_{\nu_{k+1}}$. Their respective ith elements have the weights $\hat{\mathtt{w}}_{x_k i}$, $\bar{\mathtt{w}}_{x_{k+1} i}$, $\tilde{\mathtt{w}}_{x_{k+1} i}$, $\mathtt{w}_{w_k i}$, and $\mathtt{w}_{\nu_{k+1} i}$, the mean values $\hat{\boldsymbol{\mu}}_{x_k i}$, $\bar{\boldsymbol{\mu}}_{x_{k+1} i}$, $\tilde{\boldsymbol{\mu}}_{x_{k+1} i}$, $\boldsymbol{\mu}_{w_k i}$, and $\boldsymbol{\mu}_{\nu_{k+1} i}$, and the square-root information matrices $\hat{R}_{x x_k i}$, $\bar{R}_{x x_{k+1} i}$, $\tilde{R}_{x x_{k+1} i}$, $R_{w w_k i}$, and $R_{\nu \nu_{k+1} i}$.

The Gaussian mixture $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ in Eq. (14) ostensibly represents the same distribution as the *a priori* Gaussian distribution $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ in Eq. (13), which is the output of the dynamic propagation calculation prior to re-sampling. The alternate *a priori* probability density function $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ is an approximation of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ that is generated by applying the re-sampling algorithm of Ref. [20]. The primary goal of this re-sampling algorithm is to enforce MI upper bounds on the covariance of each new mixand $\tilde{P}_{x x_{k+1} i} = \tilde{R}^{-1}_{x x_{k+1} i} \tilde{R}^{-\mathrm{T}}_{x x_{k+1} i}$. A secondary goal is to reduce the number of mixands so that $\tilde{N}_{x_{k+1}}$ is as small as possible while $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ is still a good approximation of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$.

C. Gaussian Mixtures as Approximations of General Distributions

Gaussian mixtures are used because of the ease of employing EKF or UKF operations in order to perform approximate filtering calculations on them. It is conjectured that a Gaussian mixture can accurately approximate any type of smooth probability density function if the number of mixands is large enough, if the mixand covariances are small enough, and if the means and weights of the individual mixands are chosen judiciously. This conjecture is based on the observation that mixands with small covariances tend to act like basis functions of a spline approximation scheme for a general function. Their weights and mean values can be adjusted to minimize the fit error to the original distribution if their covariances are small enough to avoid "washing out" any fine features of the original distribution. A proof of this conjecture is beyond the scope of this paper.

It should be straightforward, however, to prove that a Gaussian mixture can accurately approximate any distribution that results from applying the Bayesian dynamic propagation operations of Eq. (4) or the Bayesian measurement update operations of Eq. (7) to a distribution that was originally a Gaussian mixture. A proof of this assertion could use Taylor series and the linearization approximations of Section V.

IV. MI Covariance Bounds and Gaussian Mixture Re-Sampling

The main new feature of this paper's Gaussian mixture filter is the algorithm that generates the re-sampled Gaussian mixture $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ from $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$. The re-sampling algorithm is developed in Ref. [20]. It enforces MI upper bounds on the covariances of the individual mixands of the new $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ distribution. Enforcement of these bounds is the key to successful Bayesian estimation through the application of EKF approximations to each mixand within static multiple-model filter calculations. The present section gives an overview this re-sampling algorithm.

A. Square-Root Information MI Bounds and Candidate Square-Root Information Matrices for Re-Sampled Mixands

Given the covariance matrix upper bound $P_{\max}$, a corresponding square-root information matrix $R_{\min}$ can be computed using Cholesky factorization followed by matrix inversion so that $R^{-1}_{\min} R^{-\mathrm{T}}_{\min} = P_{\max}$. Given $R_{\min}$, the new mixands' covariances will be upper-bounded by $P_{\max}$, in the MI sense, if and only if their square-root information matrices respect the following MI lower bounds [20]:

$$\tilde{R}^{\mathrm{T}}_{x x_{k+1} i} \tilde{R}_{x x_{k+1} i} \geq R^{\mathrm{T}}_{\min} R_{\min}, \quad \text{for all } i = 1, \ldots, \tilde{N}_{x_{k+1}}. \tag{17}$$

The choice of the covariance upper bound $P_{\max}$ is problem-dependent. It must be chosen small enough so that linear approximations of the dynamics and measurement model functions in Eq. (1) are reasonably accurate over about a 2-σ range of $P_{\max}$. It may be allowable for the chosen $P_{\max}$ to permit large variations in certain subspaces where the problem model functions only have weak nonlinearities. In the limit of subspaces with perfect linearity of the model functions, very large corresponding projections of $P_{\max}$ are acceptable, and the resulting filter calculations will behave somewhat like a Rao-Blackwellized PF. It may also be allowable to make the choice of $P_{\max}$ dependent on the region of $\boldsymbol{x}_k$ space where the mixands are being generated. A convenient way to implement such a dependence is to choose $P_{\max}$ as a function of the mean value of the $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ mixand that is used to generate a particular new $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ mixand.

The re-sampling algorithm generates each new daughter mixand of $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ from an original parent mixand of

$p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$. The daughter mixand inherits its new square-root information matrix from a perturbed version of the square-root information matrix of the parent mixand. Let the perturbed square-root information matrix associated with the ith mixand of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ be called $\check{R}_{xx_{k+1}i}$. It is chosen to obey the following two MIs

$$\begin{aligned}\check{R}^{\mathrm{T}}_{xx_{k+1}i}\check{R}_{xx_{k+1}i} &\geq R^{\mathrm{T}}_{\min}R_{\min}\\ \check{R}^{\mathrm{T}}_{xx_{k+1}i}\check{R}_{xx_{k+1}i} &\geq \bar{R}^{\mathrm{T}}_{xx_{k+1}i}\bar{R}_{xx_{k+1}i}\end{aligned} \tag{18}$$

while minimizing the following weighted-matrix-norm-squared cost function:

$$J[\check{R}_{xx_{k+1}i}] = \mathrm{trace}[R^{-\mathrm{T}}_{\min}\check{R}^{\mathrm{T}}_{xx_{k+1}i}\check{R}_{xx_{k+1}i}R^{-1}_{\min}]. \tag{19}$$

Thus, $\check{R}_{xx_{k+1}i}$ is the "smallest" square-root information matrix that is no smaller, in the squared MI sense, than $R_{\min}$ or $\bar{R}_{xx_{k+1}i}$. The $R_{\min}$ MI lower bound in the first line of Eq. (18) enforces the $P_{\max}$ maximum covariance bound, and the $\bar{R}_{xx_{k+1}i}$ MI lower bound in the second line of Eq. (18) ensures that the daughter mixand covariance is also upper-bounded by the covariance of the parent mixand. This second bound is needed in order to ensure that the parent mixand can be approximated by a set of daughter mixands. Minimization of the information matrix norm-squared cost function in Eq. (19) causes the daughter mixand covariance to be as large as possible while respecting the two covariance upper bounds associated with the MIs in Eq. (18). By using the largest possible daughter covariance, the re-sampling algorithm is likely to be able to achieve good re-approximation accuracy with the fewest possible daughter mixands per parent mixand.

Section III of Ref. [20] details the calculations that solve for the $\check{R}_{xx_{k+1}i}$ which satisfies the MIs in Eq. (18) while minimizing the cost in Eq. (19). They involve a sequence of matrix linear algebra calculations that include matrix inversion, singular value decomposition, and orthonormal/upper-triangular (QR) factorization [21].

The needed calculations yield an additional matrix that is of importance to the re-sampling procedure. It is the matrix square-root of the covariance decrement in going from the parent mixand to the daughter mixand, $\delta\check{Y}_{xx_{k+1}i}$. It obeys the covariance decrement relationship

$$\begin{aligned}&\delta\check{Y}_{xx_{k+1}i}\delta\check{Y}^{\mathrm{T}}_{xx_{k+1}i}\\ &\quad= \bar{R}^{-1}_{xx_{k+1}i}\bar{R}^{-\mathrm{T}}_{xx_{k+1}i} - \check{R}^{-1}_{xx_{k+1}i}\check{R}^{-\mathrm{T}}_{xx_{k+1}i}.\end{aligned} \tag{20}$$

This covariance square-root matrix always has n_x rows, but it may have fewer than n_x columns. In fact, if $\bar{R}_{xx_{k+1}i}$ satisfies the covariance-bounding MI in the first line of Eq. (18), then the optimal $\check{R}_{xx_{k+1}i}$ equals $\bar{R}_{xx_{k+1}i}$, and $\delta\check{Y}_{xx_{k+1}i}$ can be an empty array or simply an n_x-by-1 matrix of all zeros.

B. Sampling-Based Algorithm to Choose Mixands of the New Gaussian Mixture Distribution

The weights, mean values, and square-root information matrices of the re-sampled $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ distribution are chosen by sampling from a modified version of the $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ distribution. This modified version is:

$$\begin{aligned}&p_{tempx_{k+1}}(\boldsymbol{x}_{k+1})\\ &\quad= \sum_{i=1}^{\bar{N}_{x_{k+1}}} \bar{\mathbf{w}}_{x_{k+1}i}\mathcal{N}\left[\boldsymbol{x}_{k+1};\bar{\boldsymbol{\mu}}_{x_{k+1}i}, \delta\check{Y}_{xx_{k+1}i}\delta\check{Y}^{\mathrm{T}}_{xx_{k+1}i}\right].\end{aligned} \tag{21}$$

Its definition uses the standard covariance form of the vector Gaussian distribution

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},P) = \frac{1}{(2\pi)^{n_x/2}\sqrt{\det(P)}}e^{-0.5(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}P^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{22}$$

Comparison of Eq. (13) with Eq. (21) reveals that the only difference between $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ and $p_{tempx_{k+1}}(\boldsymbol{x}_{k+1})$ is that the ith mixand of the former has covariance $\bar{R}^{-1}_{xx_{k+1}i}\bar{R}^{-\mathrm{T}}_{xx_{k+1}i}$, while the corresponding mixand of the latter has covariance $\delta\check{Y}_{xx_{k+1}i}\delta\check{Y}^{\mathrm{T}}_{xx_{k+1}i}$. The latter covariance is less than or equal to the former in the MI sense.

The re-sampling algorithm picks the means of its new mixands by sampling directly from $p_{tempx_{k+1}}(\boldsymbol{x}_{k+1})$. This sampling procedure is accomplished in two steps that use random number generators and standard Monte-Carlo techniques. The first step samples from a 1-dimensional uniform distribution on the range [0,1]. The uniform sample is compared to the cumulative weights in Eq. (21) in order to select the particular mixand of $p_{tempx_{k+1}}(\boldsymbol{x}_{k+1})$ that will be sampled to determine the mean of a new $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ mixand. Let the chosen element of the $p_{tempx_{k+1}}(\boldsymbol{x}_{k+1})$ distribution be mixand j. The second step determines the mean of the new lth mixand of $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ by using a random number generator to sample a zero-mean, identity-covariance vector Gaussian distribution with dimension equal to the number of columns in $\delta\check{Y}_{xx_{k+1}j}$. Let this random vector be $\boldsymbol{\eta}_{jl}$. The mean value of the new lth mixand of $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ becomes

$$\tilde{\boldsymbol{\mu}}_{x_{k+1}l} = \bar{\boldsymbol{\mu}}_{x_{k+1}j} + \delta\check{Y}_{xx_{k+1}j}\boldsymbol{\eta}_{jl}. \tag{23}$$

The square-root information matrix of the new lth mixand is set equal to the pre-computed value that corresponds to the MI-bounded version of the covariance of the jth mixand of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$:

$$\tilde{R}_{xx_{k+1}l} = \check{R}_{xx_{k+1}j}. \tag{24}$$

All of the weights for the new mixands are nominally equal-valued, as in standard importance re-sampling in a PF.

The basic re-sampling algorithm given above is augmented in three ways in Ref. [20]. First, it starts with a target value for the number of re-sampled mixands, N_{target}. The actual final number $\tilde{N}_{x_{k+1}}$ will be no larger than this target value, and it may be smaller. Second, it may happen that some of the mixands of $p_{tempx_{k+1}}(\boldsymbol{x}_{k+1})$ in Eq. (21) will have zero covariance, i.e., $\delta\check{Y}_{xx_{k+1}i} = 0$. If such a mixand is re-sampled multiple times by this algorithm, then a simple-minded application of the algorithm will produce multiple identical mixands of the re-sampled $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ with identical weights. Instead, the algorithm combines these mixands into a single re-sampled mixand with an appropriately increased weight. Third, several of the original mixands of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ may be nearly identical and

have $\delta\check{Y}_{xx_{k+1}i}$ values equal to zero. In this case, the re-sampling algorithm of Ref. [20] attempts to merge such mixands before executing its re-sampling steps. Merging is carried out if it results in an initial re-approximation of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ that is not very different from the original $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ as measured in a functional 2-norm sense. These last two features of the algorithm tend to limit the number of mixands in the re-sampled $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$, $\tilde{N}_{x_{k+1}}$. Experience has shown that these features are likely to reduce $\tilde{N}_{x_{k+1}}$ in situations where the underlying nonlinear/non-Gaussian Bayesian filter has converged to a sufficiently accurate solution, one whose total covariance respects the $P_{\max}$ bound.

V. Gaussian Mixture Filtering Algorithm

This paper's new Gaussian mixture filter uses the 5 mixtures in Eqs. (12)–(16) and the dynamics and measurement models in Eq. (1) to approximate the operations of an exact nonlinear/non-Gaussian Bayesian filter. The dynamic propagation operation uses $p_{x_k}(\boldsymbol{x}_k|k)$ from Eq. (12), $p_{w_k}(\boldsymbol{w}_k)$ from Eq. (15), and local linearized approximations of the dynamics model in Eq. (1) in order to compute $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ by evaluating the integral on the third line of Eq. (4). These calculations are followed by the Gaussian mixture re-sampling algorithm of Ref. [20], which has been reviewed in Section IV. It forms $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ to approximate $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ in a way that respects the MI mixand covariance upper bound which is implicitly expressed by Eq. (17). It also attempts to reduce the needed number of mixands. The final filtering step is the measurement update. It uses $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ from Eq. (14), $p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1})$ from Eq. (16), and local linearized approximations of the measurement model in Eq. (1) in order to compute $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$ by evaluating the second line of Eq. (7). The dynamic propagation calculations and the measurement update calculations are standard static multiple-model filter operations [4], except that they are carried out using EKF techniques cast in the form of SRIF calculations. It would be possible to use UKF techniques in place of EKF techniques for each mixand's calculations, but they would have to be cast in SRIF form in order to be directly usable with the present algorithm.

The next two sections define, respectively, the multiple-model/EKF/SRIF dynamic propagation and the multiple-model/EKF/SRIF measurement update. The final section combines these operations with the re-sampling algorithm of Ref. [20] and Section IV in order to define this paper's full Gaussian mixture filter. It also discusses the reason for placement of the re-sampling step between the dynamic propagation and the measurement update.

A. Gaussian Mixture Dynamic Propagation Using Mixand-by-Mixand EKF Calculations

Multiple-model filter dynamic propagation operations are used to approximate the integral in the last line of Eq. (4). The approximation uses EKF calculations for each mixand. The input mixands for this calculation are those of the product distribution:

$$p_{x_k}(\boldsymbol{x}_k|k)p_{w_k}(\boldsymbol{w}_k) = \left[\sum_{i=1}^{\hat{N}_{x_k}} \hat{\mathtt{w}}_{x_k i}\mathcal{N}_{sr}(\boldsymbol{x}_k;\hat{\boldsymbol{\mu}}_{x_k i},\hat{R}_{xx_k i})\right] \times \left[\sum_{j=1}^{N_{w_k}} \mathtt{w}_{w_k j}\mathcal{N}_{sr}(\boldsymbol{w}_k;\boldsymbol{\mu}_{w_k j},R_{ww_k j})\right]. \quad (25)$$

This product causes there to be $\bar{N}_{x_{k+1}} = \hat{N}_{x_k}N_{w_k}$ mixands in the calculation's $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ output distribution.

The multiple-model dynamic propagation implements the SRIF form of the EKF dynamic propagation steps for each of the $\bar{N}_{x_{k+1}}$ products of *a posteriori* state mixands $i = 1, \ldots, \hat{N}_{x_k}$ and process noise mixands $j = 1, \ldots, N_{w_k}$. The index of each resulting *a priori* state mixand for $\boldsymbol{x}_{k+1}$ is $l = N_{w_k}(i-1)+j$, and the corresponding dynamic propagation calculations for this mixand are:

$$\begin{aligned}\bar{\mathtt{w}}_{x_{k+1}l} &= \hat{\mathtt{w}}_{x_k i}\mathtt{w}_{w_k j}\\ \bar{\boldsymbol{\mu}}_{x_{k+1}l} &= \boldsymbol{f}_k\left(\hat{\boldsymbol{\mu}}_{x_k i},\boldsymbol{\mu}_{w_k j}\right)\\ Q_{kl}\begin{bmatrix}\hat{R}_{ww_k l} & \hat{R}_{wx_{k+1}l}\\ 0 & \bar{R}_{xx_{k+1}l}\end{bmatrix} &= \begin{bmatrix}R_{ww_k j} & 0\\ -\hat{R}_{xx_k i}\Phi_{kl}^{-1}\Gamma_{kl} & \hat{R}_{xx_k i}\Phi_{kl}^{-1}\end{bmatrix}\end{aligned} \quad (26)$$

where the calculations in the last line of Eq. (26) start with the matrix on the right-hand side and perform a QR factorization of that matrix in order to compute the outputs on the left-hand side. These outputs are the orthonormal matrix Q_{kl}, the square, upper-triangular matrices $\hat{R}_{ww_k l}$ and $\bar{R}_{xx_{k+1}l}$, and the additional matrix $\hat{R}_{wx_{k+1}l}$.

The calculations in Eq. (26) have been derived by first substituting the linearized approximation

$$\begin{aligned}\boldsymbol{x}_k &= \boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1},\boldsymbol{w}_k)\\ &\cong \hat{\boldsymbol{\mu}}_{x_k i} + \Phi_{kl}^{-1}\left\{\boldsymbol{x}_{k+1} - \boldsymbol{f}_k(\hat{\boldsymbol{\mu}}_{x_k i},\boldsymbol{\mu}_{w_k j})\right.\\ &\quad \left. -\Gamma_{kl}[\boldsymbol{w}_k - \boldsymbol{\mu}_{w_k j}]\right\}\end{aligned} \quad (27)$$

with

$$\Phi_{kl} = \left.\frac{\partial \boldsymbol{f}_k}{\partial \boldsymbol{x}_k}\right|_{(\hat{\boldsymbol{\mu}}_{x_k i},\boldsymbol{\mu}_{w_k j})} \text{ and } \Gamma_{kl} = \left.\frac{\partial \boldsymbol{f}_k}{\partial \boldsymbol{w}_k}\right|_{(\hat{\boldsymbol{\mu}}_{x_k i},\boldsymbol{\mu}_{w_k j})} \quad (28)$$

into the product term $\hat{\mathtt{w}}_{x_k i}\mathcal{N}_{sr}(\boldsymbol{x}_k;\hat{\boldsymbol{\mu}}_{x_k i},\hat{R}_{xx_k i}) \times \mathtt{w}_{w_k j}\mathcal{N}_{sr}(\boldsymbol{w}_k;\boldsymbol{\mu}_{w_k j},\hat{R}_{ww_k j})$ of the Gaussian mixture in Eq. (25). The resulting approximate term is a joint Gaussian in $\boldsymbol{x}_{k+1}$ and $\boldsymbol{w}_k$. Its integral with respect to $\boldsymbol{w}_k$ in the last line of Eq. (4) can be evaluated by using the Gaussian unit normalization condition. The values of $\bar{\mathtt{w}}_{x_{k+1}l}$, $\bar{\boldsymbol{\mu}}_{x_{k+1}l}$, and $\bar{R}_{xx_{k+1}l}$ in Eq. (26) have been defined so that the resulting integral equals $\bar{\mathtt{w}}_{x_{k+1}l}\mathcal{N}_{sr}[\boldsymbol{x}_{k+1};\bar{\boldsymbol{\mu}}_{x_{k+1}l},\bar{R}_{xx_{k+1}l}]$, consistent with standard SRIF calculations [22].

In order for the linearized approximation in Eq. (27) to be sufficiently accurate, it is required that the $p_{w_k}(\boldsymbol{w}_k)$ mixand covariances $R_{ww_k j}^{-1}R_{ww_k j}^{-\mathrm{T}}$ for $j = 1, \ldots, N_{w_k}$ be sufficiently

small. Thus, the square-root information matrices for the process noise mixands must obey an MI of the form

$$R_{ww_kj}^{\mathrm{T}} R_{ww_kj} \geq R_{wwmin}^{\mathrm{T}} R_{wwmin}, \quad \text{for all } j = 1, \ldots N_{w_k} \tag{29}$$

where $P_{wwmax} = R_{wwmin}^{-1} R_{wwmin}^{-\mathrm{T}}$ is a covariance upper bound for the *a priori* process noise mixands. As with the state mixand covariance upper bound $P_{\max}$, an appropriate choice for this bound depends on the nonlinearities in the function $\boldsymbol{f}_k^{-1}(\boldsymbol{x}_{k+1}, \boldsymbol{w}_k)$. If the original formulation of the filtering problem does not respect this process noise mixand MI, then the Gaussian mixture re-sampling algorithm of Ref. [20] must be applied to the original process noise distribution in order to produce a distribution in the form of Eq. (15) that has mixands which respect this bound. This re-sampled distribution will be used by the filter.

The orthonormal matrix Q_{kl} is discarded after the operations in Eq. (26), and the matrices $\hat{R}_{ww_kl}$ and $\hat{R}_{wx_{k+1}l}$ are also discarded unless one wants to do a backwards smoothing pass after the forwards filtering pass.

The dynamic propagation of the mixand square-root information matrix in the third line of Eq. (26) relies on the inverse of the mixand state transition matrix Φ_{kl}. Alternate calculations can be implemented in situations where Φ_{kl} is not invertible, provided that the resulting $\bar{R}_{xx_{k+1}l}$ will not be infinite. This condition will be met for any process noise model that keeps the filtering problem non-singular. One suitable alternate version uses square-root covariance propagation followed by matrix inversion in order to revert back to a square-root information matrix representation of the mixand covariance.

B. Gaussian Mixture Measurement Update Using Mixand-by-Mixand EKF Computations

A multiple-model filter measurement update is used in conjunction with EKF techniques in order to approximate the last line of Eq. (7). The input mixands for this calculation are those of the product distribution:

$$\begin{aligned}
&\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k) p_{\nu_{k+1}}\left[\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})\right] \\
&\quad = \left(\sum_{i=1}^{\tilde{N}_{x_{k+1}}} \tilde{\mathtt{w}}_{x_{k+1}i} \mathcal{N}_{sr}\left[\boldsymbol{x}_{k+1}; \tilde{\boldsymbol{\mu}}_{x_{k+1}i}, \tilde{R}_{xx_{k+1}i}\right] \right) \\
&\quad \times \left(\sum_{j=1}^{N_{\nu_{k+1}}} \mathtt{w}_{\nu_{k+1}j} \mathcal{N}_{sr}\left\{ \left[\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})\right]; \right.\right. \\
&\qquad \left.\left. \boldsymbol{\mu}_{\nu_{k+1}j}, R_{\nu\nu_{k+1}j} \right\} \right).
\end{aligned} \tag{30}$$

Thus, there are $\hat{N}_{x_{k+1}} = \tilde{N}_{x_{k+1}} N_{\nu_{k+1}}$ mixands in the filter's final *a posteriori* distribution $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$.

The EKF/SRIF multiple-model measurement update performs the Bayesian calculations in the last line of Eq. (7) using local linearizations of the measurement model function $\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})$. The linearized approximation used in the ijth product term of Eq. (30) is:

$$\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1}) \cong \boldsymbol{h}_{k+1}\left[\tilde{\boldsymbol{\mu}}_{x_{k+1}i}\right] + H_{(k+1)i}\left[\boldsymbol{x}_{k+1} - \tilde{\boldsymbol{\mu}}_{x_{k+1}i}\right] \tag{31}$$

where

$$H_{(k+1)i} = \left.\frac{\partial \boldsymbol{h}_{k+1}}{\partial \boldsymbol{x}_{k+1}}\right|_{\tilde{\boldsymbol{\mu}}_{x_{k+1}i}}. \tag{32}$$

The measurement update re-arranges each mixand product term in Eq. (30), $\tilde{\mathtt{w}}_{x_{k+1}i}\mathcal{N}_{sr}[\boldsymbol{x}_{k+1}; \tilde{\boldsymbol{\mu}}_{x_{k+1}i}, \tilde{R}_{xx_{k+1}i}] \times \mathtt{w}_{\nu_{k+1}j}\mathcal{N}_{sr}\{[\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})]; \boldsymbol{\mu}_{\nu_{k+1}j}, R_{\nu\nu_{k+1}j}\}$, into the equivalent *a posteriori* form $\hat{\mathtt{w}}_{unx_{k+1}l}\mathcal{N}_{sr}[\boldsymbol{x}_{k+1}; \hat{\boldsymbol{\mu}}_{x_{k+1}l}, \hat{R}_{xx_{k+1}l}]$. This is done for each of the $\hat{N}_{x_{k+1}}$ products of *a priori* state mixands $i = 1, \ldots, \tilde{N}_{x_{k+1}}$ and measurement noise mixands $j = 1, \ldots, N_{\nu_{k+1}}$. The index of each corresponding *a posteriori* state mixand is $l = N_{\nu_{k+1}}(i-1) + j$. Equivalence between the two forms is possible because substitution of the linearized measurement function model from Eq. (31) into the measurement noise mixand causes the original product term to be Gaussian in $\boldsymbol{x}_{k+1}$. The following standard SRIF measurement update calculations [22] and multiple-model filter calculations are designed to ensure the equivalence of these two Gaussian forms

$$\hat{Q}_{(k+1)l}\begin{bmatrix} \hat{R}_{xx_{k+1}l} \\ 0 \end{bmatrix} = \begin{bmatrix} \tilde{R}_{xx_{k+1}i} \\ R_{\nu\nu_{k+1}j} H_{(k+1)i} \end{bmatrix} \tag{33}$$

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{z}}_{x_{k+1}l} \\ \boldsymbol{z}_{r_{k+1}l} \end{bmatrix} &= \hat{Q}_{(k+1)l}^{\mathrm{T}} \\
&\times \begin{bmatrix} 0 \\ R_{\nu\nu_{k+1}j}\left\{\boldsymbol{y}_{k+1} - \boldsymbol{h}_{k+1}(\tilde{\boldsymbol{\mu}}_{x_{k+1}i}) - \boldsymbol{\mu}_{\nu_{k+1}j}\right\} \end{bmatrix}
\end{aligned} \tag{34}$$

$$\hat{\boldsymbol{\mu}}_{x_{k+1}l} = \tilde{\boldsymbol{\mu}}_{x_{k+1}i} + \hat{R}_{xx_{k+1}l}^{-1} \hat{\boldsymbol{z}}_{x_{k+1}l} \tag{35}$$

$$\begin{aligned}
\hat{\mathtt{w}}_{unx_{k+1}l} &= \tilde{\mathtt{w}}_{x_{k+1}i}\mathtt{w}_{\nu_{k+1}j} e^{-0.5\boldsymbol{z}_{r_{k+1}l}^{\mathrm{T}}\boldsymbol{z}_{r_{k+1}l}} \\
&\times \left\{ \frac{|\det[\tilde{R}_{xx_{k+1}i}]\det[R_{\nu\nu_{k+1}j}]|}{|\det[\hat{R}_{xx_{k+1}l}]|(2\pi)^{n_y/2}} \right\}.
\end{aligned} \tag{36}$$

The calculations in Eq. (33) QR factorize the input matrix on the right-hand side in order to compute the orthonormal matrix $\hat{Q}_{(k+1)l}$ and the square, upper-triangular *a posteriori* square-root information matrix $\hat{R}_{xx_{k+1}l}$, which both appear on the left-hand side. Equation (34) forms the normalized measurement residual on the right-hand side and transforms the result using the transpose of $\hat{Q}_{(k+1)l}$ in order to compute the state update vector $\hat{\boldsymbol{z}}_{x_{k+1}l}$ and the residual error vector $\boldsymbol{z}_{r_{k+1}l}$. The former vector is used in Eq. (35) to update the mixand mean.

Equation (36) computes the un-normalized version of the *a posteriori* mixand weight. Four of the product terms on the right-hand side of Eq. (36) are standard in a static multiple-model re-weighting formula: the two pre-update mixand weights, the power of 2π, and the exponential that involves a sum of squares of normalized measurement residuals [4]. The terms involving determinants of square-root information matrices are required

for this new SRIF form of the static multiple-model measurement update. They replace a term involving the determinant of the innovations covariance matrix that appears in the covariance version of this re-weighting formula [4].

The final step of the Gaussian mixture measurement update re-normalizes the new mixand weights. Re-normalization occurs after the calculations in Eqs. (33)–(36) have been carried out for all of the mixands. The re-normalization takes the form:

$$\hat{\mathtt{w}}_{x_{k+1}l} = \frac{\hat{\mathtt{w}}_{unx_{k+1}l}}{\sum_{m=1}^{\hat{N}_{x_{k+1}}} \hat{\mathtt{w}}_{unx_{k+1}m}} \text{ for } l = 1, \ldots, \hat{N}_{x_{k+1}}. \quad (37)$$

This completes the calculations on the last line of Eq. (7) because the sum of the unnormalized weights in the denominator of Eq. (37) equals the integral of the probability density product in the denominator of Eq. (7).

Given that these calculations could involve a large number of mixands, $\hat{N}_{x_{k+1}}$, it is important to carry them out in the most efficient manner possible. One important efficiency is to evaluate the linearization calculations in Eqs. (31) and (32) only $\tilde{N}_{x_{k+1}}$ times, not $\hat{N}_{x_{k+1}}$ times. Similarly, $\det[\tilde{R}_{xx_{k+1}i}]$ only needs to be evaluated $\tilde{N}_{x_{k+1}}$ different times, and $\det[R_{\nu\nu_{k+1}j}]$ only requires $N_{\nu_{k+1}}$ independent evaluations. Furthermore, the usual upper-triangularity of $\tilde{R}_{xx_{k+1}i}$, $R_{\nu\nu_{k+1}j}$, and $\hat{R}_{xx_{k+1}l}$ expedites the determinant calculations because the determinant of an upper-triangular matrix equals the product of its diagonal elements.

C. Complete Gaussian Mixture Filter Operations

The Gaussian mixture filtering algorithm starts with the following inputs:

1) the *a posteriori* state Gaussian mixture at sample 0, $p_{x_0}(\boldsymbol{x}_0|0)$,
2) the process noise and measurement noise Gaussian mixtures $p_{w_k}(\boldsymbol{w}_k)$, and $p_{\nu_{k+1}}(\boldsymbol{\nu}_{k+1})$, at samples $k = 0, 1, 2, \ldots$
3) the square-root information matrix MI lower bound $R_{\min}$ that is used in Eq. (17), and
4) the target number of mixands after each state re-sampling, N_{target}.

Given these inputs, the approximate nonlinear/non-Gaussian Bayesian calculations of the new Gaussian mixture filter proceed according to the following 7 steps:

1) If necessary, re-sample $p_{x_0}(\boldsymbol{x}_0|0)$ using the Gaussian mixture re-sampling algorithm of Ref. [20] in order to ensure that the MI in Eq. (17) is satisfied by each mixand's square-root information matrix. If re-sampling is necessary, then use N_{target} as the target/maximum number of re-sampled mixands.
2) Set the current sample index to $k = 0$.
3) Perform the multiple-model SRIF/EKF dynamic propagation calculations in Eq. (26) for all $\bar{N}_{x_{k+1}} = \hat{N}_{x_k} N_{w_k}$ mixands of the new *a priori* distribution $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$. Use the matrices Φ_{kl} and Γ_{kl} for $l = 1, \ldots, \bar{N}_{x_{k+1}}$ from Eq. (28) in these calculations.
4) Perform Gaussian mixture re-sampling of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ in order to produce $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ with mixand square-root information matrices that all satisfy the MI in Eq. (17). Use the re-sampling procedure of Ref. [20], as discussed in this paper's Section IV. Use N_{target} as the target/maximum number of re-sampled mixands.
5) Perform the multiple-model SRIF/EKF measurement update calculations in Eqs. (33)–(36) for all $\hat{N}_{x_{k+1}} = \tilde{N}_{x_{k+1}} N_{\nu_{k+1}}$ mixands of the new *a posteriori* distribution $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$. Use the matrices $H_{(k+1)i}$ for $i = 1, \ldots, \tilde{N}_{x_{k+1}}$ from Eq. (32) in these calculations.
6) Use Eq. (37) to compute the normalized mixand weights of the new *a posteriori* distribution $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$.
7) Replace k by $k+1$. Stop if the incremented k value is the last sample of the filtering run. Otherwise, return to Step 3.

This is a recursive procedure. Each iteration of Steps 3–7 produces the *a posteriori* distribution $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k+1)$ that becomes the input distribution $p_{x_k}(\boldsymbol{x}_k|k)$ for the next iteration after k is incremented in Step 7.

The most important purpose of Step 4's mixture re-sampling is to bound the mixand covariances using the MI in Eq. (17). This bounding ensures sufficient accuracy of the EKF linearizations of the dynamics and measurement models in Eqs. (27) and (31). It is obvious that enforcement of the MI bound in Step 4 can ensure accuracy of the measurement model linearization in Eq. (31) because the measurement update calculations of Step 5 occur immediately after the re-sampling step. Satisfaction of the MI bound in Eq. (17) during the dynamic propagation may seem uncertain because the measurement update in Steps 5 and 6 happens after the re-sampling and before the Step-3 dynamic propagation of the filter recursion's next iteration. Note, however, that the SRIF measurement update calculation in Eq. (33) guarantees that

$$\hat{R}^{\mathrm{T}}_{xx_{k+1}l}\hat{R}_{xx_{k+1}l} \geq \tilde{R}^{\mathrm{T}}_{xx_{k+1}i}\tilde{R}_{xx_{k+1}i}. \quad (38)$$

This inequality re-states the fact that the measurement update of a linear filter can never increase the covariance of the corresponding state estimate. The re-sampling in algorithm Step 4 ensures that the MI in Eq. (17) is satisfied by each $\tilde{R}_{xx_{k+1}i}$. Equation (38) and the transitivity of MIs ensure that

$$\hat{R}^{\mathrm{T}}_{xx_{k+1}l}\hat{R}_{xx_{k+1}l} \geq R^{\mathrm{T}}_{\min}R_{\min}, \quad \text{for all } l = 1, \ldots, \hat{N}_{x_{k+1}}. \quad (39)$$

Therefore, the filter need not re-sample after its measurement update. Equation (39) guarantees that the *a posteriori* mixands at the end of the kth iteration of Steps 3–7 are sufficiently narrow to admit accurate EKF dynamic propagation calculations in Step 3 of the $k+1$st iteration.

The second goal of the mixture re-sampling in Step 4 is to limit the number of mixands. If possible, this step seeks to reduce this number as much as is practical without adversely affecting the accuracy of the re-sampled distribution. One can show that $\bar{N}_{x_{k+1}} = \tilde{N}_{x_k} N_{\nu_k} N_{w_k}$. If N_{ν_k} or N_{w_k} is greater than 1, then the number of mixands will grow during execution of Steps 5–7 of the $k-1$st iteration of the algorithm and Step 3 of the kth iteration. Without the re-sampling in Step 4, the number of mixands would grow geometrically without bound, and the calculations would quickly become intractable. The re-sampling in Step 4

bounds each $\tilde{N}_{x_k}$ to be no greater than N_{target}, thereby bounding the possible growth in the number of mixands. Normally such a bound can be enforced without undue loss of accuracy because many of the mixands "lost" in the re-sampling process have insignificant weights or are redundant with other mixands.

As will be shown in Section VI, the re-sampling algorithm of Ref. [20] includes features that have the potential to form a $\tilde{p}_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$ that is an accurate approximation of $p_{x_{k+1}}(\boldsymbol{x}_{k+1}|k)$, but with the number of re-sampled mixands $\tilde{N}_{x_{k+1}}$ much less than N_{target}. When this happens, the Gaussian mixture filter has the potential to be very efficient computationally, almost as efficient as a standard EKF or UKF. In other cases, however, $\tilde{N}_{x_{k+1}}$ can remain at values as large as 10,000 or even higher. In such cases, the filter computations will be 10,000 or more times as expensive per dynamic propagation and update step as an EKF or UKF. Fortunately, these periods of high computational cost are often transient, as will be seen in Section VI.

The new Gaussian mixture filter differs markedly from a standard PF in its placement of the re-sampling algorithm between the dynamic propagation of Step 3 and the measurement update of Steps 5 and 6. A typical PF performs importance re-sampling after the measurement update [5], [9]. The re-positioning of the re-sampling step in the present filter has been driven by its need to enforce the MI bound in Eq. (17). If the re-sampling step took place after the measurement update, then the potential for mixand covariance growth during the dynamic propagation on the next iteration could cause violation of the MI in Eq. (17) at the point of measurement model linearization in Eqs. (31) and (32). The fact of covariance shrinkage in the measurement update allows the filter to avoid all such problems by placing the re-sampling step between the dynamic propagation and the measurement update. This placement enables each filter iteration's one re-sampling operation to enforce the MI bound in Eq. (17) at two separate points of the iteration.

The only negative consequence of moving the re-sampling step is the potential to retain insignificant mixands too long. The re-weighting calculations of the measurement update in Eqs. (36) and (37) can cause some updated mixands to have negligible weights. Typically this happens when a given mixand has unreasonably large normalized measurement residuals in its $\boldsymbol{z}_{r_{k+1}l}$ vector from Eq. (34). Such mixands have virtually no impact on the remainder of the filter calculations. Failure to eliminate them through importance re-sampling after the measurement update results in wasted computation. The wasted operations, however, persist only until re-sampling Step 4 on the next iteration of the algorithm. That step includes mixand importance considerations in its re-sampling calculations, which eliminate all mixands with negligible weights.

VI. Gaussian Mixture Filter Performance on the Blind Tricyclist Problem

Monte-Carlo simulation tests of the new Gaussian mixture filter have been carried out for the Blind Tricyclist nonlinear filtering problem of Ref. [17]. Its performance is compared to that of the same nonlinear filters that are evaluated in Ref. [17], an EKF, two UKFs, two moving horizon estimators (also called BSEKFs), and two PFs.

A. Review of Blind Tricyclist Problem

The Blind Tricyclist Problem is developed in Ref. [17] as a benchmark nonlinear filtering problem. It involves a blind tricyclist who navigates around an amusement park based on relative bearings measured between his heading and the locations of 2 friends on 2 merry-go-rounds who shout to him intermittently. He knows which friend is on which merry-go-round and can distinguish their voices. He also knows the location and diameter of each merry-go-round, but he does not know each friend's initial phase angle on the merry-go-round (i.e., which horse the friend rides), nor does he know the constant rate of rotation of each merry-go-round. He seeks to determine his east-west/north-south position and his heading angle. In order to do this, he also must estimate the "nuisance" parameters that characterize the two friends' merry-go-rounds' instantaneous angles and constant rates. Thus, his state vector $\boldsymbol{x}_k$ has 7 elements, his position coordinates X_k and Y_k, his heading angle θ_k, the phase angles of the two friends' merry-go-round positions ϕ_{1k} and ϕ_{2k}, and their merry-go-rounds' constant rates $\dot{\phi}_{1k}$ and $\dot{\phi}_{2k}$. The k subscript on the latter two quantities is somewhat misleading in that their true values do not change with time, but their estimates can change with time as a result of measurement updates.

The blind tricyclist uses a kinematic model that includes known input time histories for the steer angle and the speed. The process noise includes small white-noise errors in these two commanded input time histories along with wheel slippage that allows slight violations of the tricycle's planar rolling constraints. The dynamics models for the merry-go-rounds are kinematic and have no uncertainty beyond their unknown initial angles and rates.

Relative bearing measurements to each merry-go-round-riding friend are made once every 3 seconds, with the two friends shouting out of phase with each other. During the 141 second duration of the considered filtering run, there are a total of 47 relative bearing measurements made to each friend. The filter sample intervals are only 0.5 seconds long. Thus, there are two sample times with no measurement data between each sample time that has bearing data. The filter algorithm skips measurement update Steps 5 and 6 in the absence of bearing data.

Fig. 1 illustrates this estimation problem. It shows a top view of the amusement park with the two merry-go-rounds and a simulated true tricyclist trajectory for a particular case. Also shown are various filters' estimates of this trajectory. The merry-go-rounds lie slightly to the left of center, with a smaller one to the north and a larger one to the south, both drawn in light grey. The truth tricyclist trajectory is shown in solid black. The tricyclist starts at the black asterisk position to the south-east of the figure's center. He proceeds almost due north until he takes a $\sim 90^\circ$ right turn. At the end of this first turn he stops and then executes a $\sim 90^\circ$ left turn traveling in reverse. At the end of this second turn he stops again at the northern-most point of his trajectory. He is now facing south. Next, he proceeds south until he is almost due east of the mid-point between the two

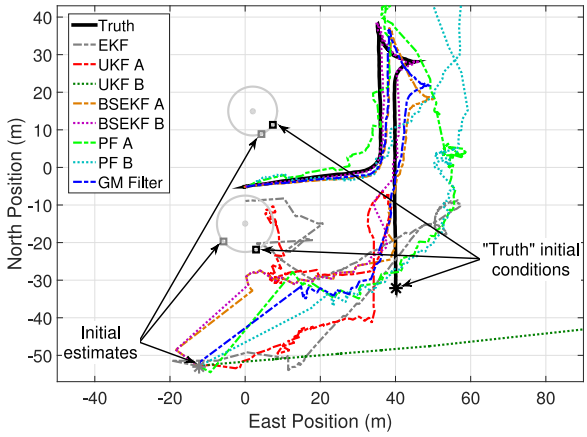

Fig. 1. Plan view of the blind tricyclist problem showing the two merry-go-rounds, the simulated truth position trajectory (black solid curve), and the corresponding estimated trajectories for 8 different filters.

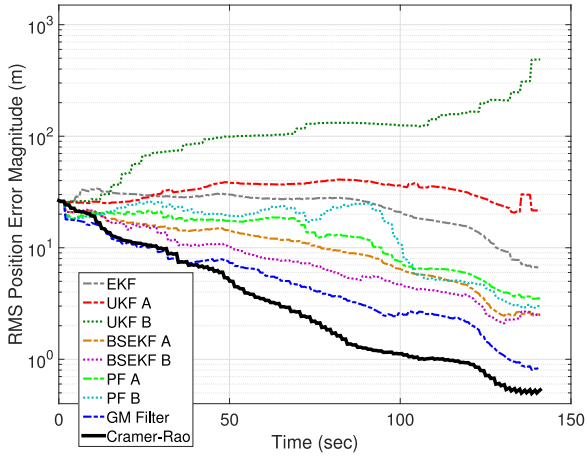

Fig. 2. Blind tricyclist RMS position error magnitude time histories for 8 different filters and the corresponding Cramer-Rao lower bound, as computed using a 100-case Monte-Carlo simulation.

merry-go-rounds. He now executes a second $\sim 90^\circ$ right turn and heads almost due west until he finishes his travels at a point between the two merry-go-rounds.

Many additional details about the blind tricyclist are contained in Ref. [17]. They include the equations that define the dynamics function $\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$ and the measurement function $\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})$, the process-noise and measurement-noise covariances, the tricycle geometry, the merry-go-round geometry, and the initial estimation error covariance. Two problems are considered, one with a small initial covariance that allows a simple EKF to converge reasonably well, and another with a much larger initial uncertainty that causes severe convergence problems for a number of the filters. Only the difficult latter case is considered here. Reference [17] cites a link to MATLAB code that includes all Blind Tricyclist model functions, a truth-model simulation, and example commands that implement an EKF solution. That link is included here as Ref. [23].

B. Comparison of the New Gaussian Mixture Filter With Other Nonlinear Filters for a Representative Blind Tricyclist Case

The Gaussian mixture filter has been applied to a truth-model simulation of the Blind Tricyclist Problem. This implementation uses $N_{w_k} = 1$ mixand for the process noise and $N_{\nu_{k+1}} = 1$ mixand for the measurement noise, thereby modeling these as Gaussian white-noise sequences. There is no need to re-sample the process noise 1-element Gaussian "mixture" in order to respect the bound on its covariance associated with Eq. (29) because its original covariance already respects a sufficiently small bound. It would be interesting to test the new filter's purported ability to handle process noise and measurement noise that were non-Gaussian and that had to be approximated by mixtures with N_{w_k} and $N_{\nu_{k+1}} > 1$, but such an investigation has been left for a future study.

The Gaussian mixture filter's re-sampling algorithm has been tuned for the Blind Tricyclist problem as follows: The value of N_{target} is set to 7,000 re-sampled mixands, though results will show that the filter is able to reduce the number of mixands far below this value after initial convergence has been achieved. The MI lower-bound on the state mixand square-root information matrices in Eq. (17) is enforced using the value $R_{\min} = \text{diag}[1/(2.6\text{ m}); 1/(2.6\text{ m}); 1/(1.04\text{ rad}); 1/(0.3467\text{ rad}); 1/(0.4\text{ rad}); 1/(2000\text{ rad/sec}); 1/(2000\text{ rad/sec})]$. The first two values correspond to maximum east-west and north-south position error standard deviations of 2.6 m for each mixand. The third term implements a maximum heading standard deviation per mixand of 1.04 rad (60°). The fourth and fifth terms specify each mixand's maximum angular standard deviations for the two merry-go-round phase angles, 0.3467 rad (20°) for the southern merry-go-round and 0.4 rad (23°) for the northern one. The last two terms are per-mixand merry-go-round rate standard-deviation limits of 2000 rad/sec (318 Hz or 1.15×10^5 °/sec). These rate limits are intentionally very large because the rates only enter linearly into the dynamics function $\boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$ and not at all into the measurement function $\boldsymbol{h}_{k+1}(\boldsymbol{x}_{k+1})$. Their nonlinear effects are indirect, through the merry-go-round phase angle terms. The mixand covariance limits on the phase angles suffice to limit the indirect nonlinear effects of large merry-go-round rate variances.

Fig. 1 plots the east-west and north-south tracking performance of the Gaussian mixture filter on an example simulated problem. The initial position estimate of the filter is given by the grey asterisk in the lower left-hand part of the figure. This is obviously far from the initial truth position designated by the black asterisk in the bottom half of the figure and somewhat right of center. Also shown in the figure are the estimated position time histories of 7 other filters that have been tested in Ref. [17]. In fact, Fig. 1 corresponds to the case that produced Fig. 4 of Ref. [17]. Therefore, most of the other filters' estimated trajectories in Fig. 1 are identical to those of Ref. [17]'s Fig. 4.

For the two PFs, however, the trajectories in Fig. 1 differ from those of Ref. [17]. During the course of the present research, it was discovered that PFs and Gaussian mixture filters can be confused by the 2π cycle ambiguities of the tricycle heading angle θ_k and the merry-go-round phase angles ϕ_{1k} and ϕ_{2k}. In the Gaussian mixture filter, this aliasing-like effect is benign. It only impacts the filter's computation of the overall state mean

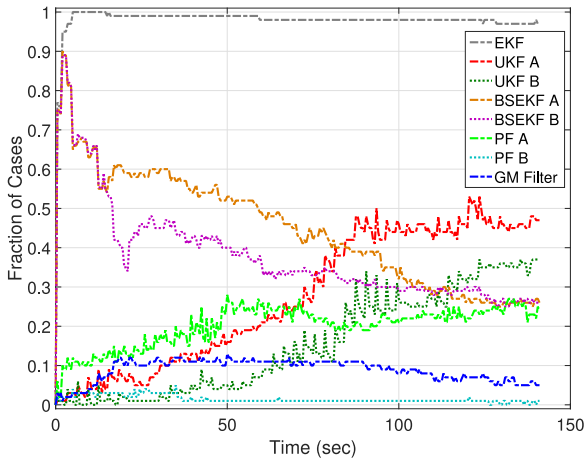

Fig. 3. Blind tricyclist consistency tests for 8 different filters that show the percentage of normalized squared state errors that lie above the 99.99% threshold of a degree-7 χ^2 statistic, as computed using 100 Monte-Carlo simulation cases.

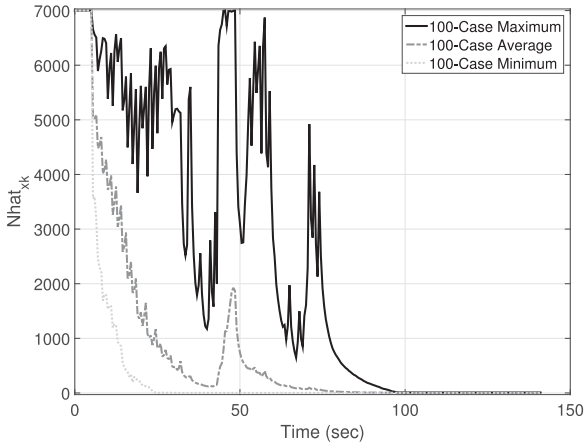

Fig. 4. Time histories of the maximum, mean, and minimum statistics of the Gaussian mixture filter's *a posteriori* mixand count $\hat{N}_k$ for 100 Monte-Carlo simulation cases.

and error covariance, as per Eq. (11). If this mean and covariance are required as outputs, then one can remove this aliasing effect by performing a 2π relative unwrapping operation between the mean heading angles and the mean merry-go-round phase angles of the filter's individual mixands. For the PFs, however, failure to compensate for this effect during filter operation can have a dramatic impact on the regularization process that has been used to avoid collapse of particle diversity. It is necessary to do inter-particle unwrapping of these angles after each measurement update before regularized re-sampling occurs. The results reported in Ref. [17] do not include this type of PF data processing. The results reported here include it. Therefore, the PF results in the present paper exhibit dramatic improvements in comparison to those of Ref. [17].

Even with the improved PF results, the best three filters in Fig. 1 are BSEKF A (light brown dash-dotted curve), BSEKF B (purple dotted curve), and the new Gaussian mixture filter (blue dash-dotted curve). These are the only filters that track near the truth black solid curve during its initial north-bound leg. They are the only filters that show the initial two turns with the stops at the end of each and with the backing-up maneuver performed during the second turn. The EKF (grey dash-dotted curve) and UKF A (red dash-dotted curve) never produce anything like the true trajectory, though they do manage to stay within the general area of tricycle motion. UKF B (dark green dotted curve) diverges out of the figure's field of view and never returns. Its sigma points spread is tuned somewhat differently than that of UKF A [17]. PF A (light green dash-dotted curve) uses 3,000 particles, and PF B (cyan dotted curve) uses 10,000. Neither of them reproduces the stopping or left-turning back-up maneuver, and PF B even exits the figure's field of view temporarily. In their favor, both PFs eventually approach the true trajectory and the estimated trajectories of the 3 good filters on the final westward leg. The best performance in Fig. 1 appears to be that of BSEKF B, which employs explicit nonlinear smoothing for the 40 half-second sample/propagation intervals that precede the filter sample of interest. The next best performance appears to be that of the new Gaussian mixture filter, and BSEKF A has the 3rd best performance. BSEKF A employs explicit nonlinear smoothing over a shorter window, only 30 half-second sample/propagation intervals. The performance of the new filter is only slightly better than that of BSEKF A, and this better performance occurs primarily during the first right-ward turn, the subsequent stop, and the initial part of the backwards left-hand turn.

C. Statistical Filter Analysis Using 100 Monte-Carlo Simulation Cases

A Monte-Carlo study has been conducted using 100 different truth-model simulations. Each simulation had the same nominal starting point for the tricyclist and the same nominal steer and speed time history as produced the black solid curve in Fig. 1. Each truth trajectory varied somewhat from that in Fig. 1 due to process noise. The filter's initial distribution $p_{x_0}(\boldsymbol{x}_0|0)$ was a Gaussian with the same large covariance for each case, but with a different mean value for each of the 100 Monte-Carlo runs. The distribution of mean values was chosen to cause the 100-case statistics of the actual errors between the truth initial state and each $p_{x_0}(\boldsymbol{x}_0|0)$ mean to be consistent with the common covariance of the 100 $p_{x_0}(\boldsymbol{x}_0|0)$ distributions.

Figs. 2 and 3 present summary statistical results from these simulations. They compare the new Gaussian mixture filter with the 7 other filters that have been used to generate Fig. 1. Fig. 2 plots the root-mean-square (RMS) east-west/north-south position error magnitude time history for each filter, as averaged over the 100 Monte-Carlo cases. It also plots the Cramer-Rao lower bound for this error as approximated by averaging over the 100 simulations. The better filter is the one with the lowest RMS error in Fig. 2. Fig. 3 plots the fraction of the 100 cases in which the normalized square of the state error $(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^{\mathrm{T}} \hat{P}_{xx_k}^{-1} (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)$ exceeds the 99.99% probability threshold of a degree-7 χ^2 distribution. In this calculation, $\hat{\boldsymbol{x}}_k$ and $\hat{P}_{xx_k}$ are, respectively, the mean and covariance of a given filter's approximation of the *a posteriori* probability density function $p_{x_k}(\boldsymbol{x}_k|k)$. The fraction of cases that violate this limit should only be about 0.0001 if the state estimation error is nearly Gaussian. Therefore, a consistent

filter will produce a curve at or very near 0 in Fig. 3. Note that Figs. 2 and 3 are nearly identical to Figs. 5 and 6 of Ref. [17], except that PF A and PF B show better performance than in Ref. [17] due to the fixed 2π ambiguity problem noted earlier.

It is obvious from Fig. 2 that the new Gaussian mixture filter performs significantly better than any of the other filters. Its RMS position error time history (blue dash-dotted curve) is consistently the lowest, and it is much closer to the Cramer-Rao lower bound (black solid curve) than that of the next best filter, BSEKF B (purple dotted curve). Thus, the superior performance of BSEKF B in Fig. 1 appears to be an outlier, a statistical exception to the rule of Fig. 2 that the new filter achieves the best RMS accuracy throughout the trajectory.

PF B (cyan dotted curve) displays the best the filter consistency in Fig. 3. Its curve is always very close to zero. The new Gaussian mixture filter (blue dash-dotted curve) has the second best consistency after $t = 70$ seconds, but one or two other filters display better consistency in the early stages of the filtering interval, UKF A (red dash-dotted curve) and UKF B (dark green dotted curve). Given that PF B, UKF A, and UKF B are significantly less accurate than the new filter, as demonstrated in Fig. 2, their superior consistency is not very useful. The two filters with accuracy closest to that of the new filter in Fig. 2 are BSEKF A (light brown dash-dotted curve) and BSEKF B (purple dotted curve). In Fig. 3, however, both of them display significantly poorer consistency than the new filter. In summary, the new Gaussian mixture filter achieves the best accuracy for this problem, and it's consistency is better than 4 of the other filters for the entire filtering interval and better than 6 of the 7 other filters during the second half of the interval.

Fig. 4 investigates the mixand count performance of the Gaussian mixture filter for the 100 Monte-Carlo simulation cases. It plots max($\hat{N}_{x_k}$) versus t_k (black solid curve), mean($\hat{N}_{x_k}$) versus t_k (grey dash-dotted curve), and min($\hat{N}_{x_k}$) versus t_k (light grey dotted curve), where the maximum, mean, and minimum are computed over the 100 Monte-Carlo cases. These curves all start at the value 7,000, which is the N_{target} value that has been used for the re-sampling in Step 4 of Section V-C. Eventually, the re-sampling algorithm is able to reduce this number through its two *ad hoc* strategies for developing an accurate Gaussian mixture re-approximation with a reduced number of mixands [20]. By the end of the filter run, the maximum $\hat{N}_{x_k}$ is 8, the mean is 3.77, and the minimum is 1.

In a PF context, the low terminal values of $\hat{N}_{x_k}$ would be interpreted as a collapse of particle diversity and would indicate filter failure. In the Gaussian mixture context, the filter can function well with a very small number of mixands if the true *a posteriori* probability density function is well modeled by the resulting mixture. Clearly this is the case for the present example, as indicated by the excellent accuracy results given in Fig. 2.

Note that the Gaussian mixture filter has the ability to increase $\hat{N}_{x_k}$ should an increase be needed. This ability relies on a coupling of two processes. The first process is the expansion of the mixand covariances of the state probability distribution that can occur during dynamic propagation. The second process is the enforcement of the MI in Eq. (17) during Gaussian mixture re-sampling. A filter with few mixands could experience mixand covariance expansion during the dynamic propagation to the point of violating the MI in Eq. (17) after the propagation. The re-sampling procedure would split the expanded mixands into multiple mixands. This expansion and splitting procedure could continue for multiple samples and eventually result in a large number of mixands. This scenario has occurred in simulation tests of the new Gaussian mixture filter on a 6-state satellite orbit determination problem.

One last filter performance metric to consider is the needed computational resources. The average time to run each filter over the 141 second data interval has been computed, with averaging carried out over the 100 Monte-Carlo cases. These averages have been computed when running the filter algorithms on a 3 GHz Windows XP Professional Workstation using MATLAB code. The results are as follows: EKF 0.08 sec, UKFs A and B 1.18 sec, BSEKF A 60.84 sec, BSEKF B 110.6 sec, PF A 149 sec, PF B 695 sec, and the new Gaussian mixture filter 187 sec. Thus, the new filter is the second most expensive filter, but it uses only 27% as much computation time, on average, as does the 10,000-particle PF B. Note, also, that the new filter is only 70% more expensive computationally than the next most accurate filter, BSEKF B. Furthermore, the new filter would have a more favorable average time comparison for longer filter runs because its largest computational costs occur during the convergence from the large initial errors, when it needs to use many mixands in order to accurately approximate the underlying probability density functions, as documented in Fig. 4.

The new filter is highly parallelizable. The only calculations that require "communication" between mixands are the weight normalization at the end of the measurement update in Eq. (37) and the parts of the re-sampling algorithm that merge mixands. Therefore, greatly increased execution speed could be achieved by mapping the algorithm onto a parallel processor.

VII. CONCLUSION

A new Gaussian mixture nonlinear filter has been developed. It uses EKF calculations to implement a static multiple-model filter in which each of its Gaussian mixands constitutes a model. These calculations are implemented in SRIF form. The key new element of this filter is its re-sampling algorithm, which executes between the dynamic propagation step and the measurement update step. The primary goal of the re-sampling algorithm is to produce an accurate approximation of the original *a priori* distribution while enforcing a MI upper bound on the covariance of each of its mixands. If this bound is tuned properly for a given problem, then it ensures that the multiple-model EKF computations will accurately approximate the underlying Bayesian calculations of an exact nonlinear/non-Gaussian filter. The re-sampling step also limits the number of mixands in the distribution, and it can reduce the number of mixands significantly when many of them would otherwise be redundant.

The new filter has been tested on a difficult 7-state benchmark nonlinear filtering problem, the Blind Tricyclist Problem. Monte-Carlo simulation tests demonstrate that the new filter is more accurate than a number of other filters, including an

EKF, two UKFs, two Moving-Horizon Estimators/BSEKFs, and two regularized PFs. Its accuracy is significantly closer to the Cramer-Rao lower bound than that of the two next best filters, the BSEKFs. Its speed of execution is slow, but it requires only 27% as much computational time as the most expensive filter considered in this study, a 10,000-particle PF. The new filter's consistency between its computed and actual estimation error covariance is imperfect, but not nearly as imperfect as the two next best filters. Given the new filter's superior accuracy, its bounded computational costs, and its reasonable consistency, it represents a good candidate for solution of difficult nonlinear filtering problems.

REFERENCES

[1] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng.*, vol. 82, pp. 34–45, Mar. 1960.

[2] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME J. Basic Eng.*, vol. 83, pp. 95–108, Mar. 1961.

[3] R. Brown and P. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, 3rd ed. New York, NY, USA: Wiley, 1997.

[4] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001.

[5] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter*. Boston, MA, USA: Artech House, 2004.

[6] F. Daum, "Nonlinear filters: Beyond the Kalman filter," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, pp. 57–69, Aug. 2005.

[7] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[8] E. Wan and R. van der Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. New York, NY, USA: Wiley, 2001, pp. 221–280.

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

[10] C. Rao, J. Rawlings, and D. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE. Trans. Autom. Control*, vol. 48, no. 2, pp. 246–258, Feb. 2003.

[11] M. Psiaki, "Backward-smoothing extended Kalman filter," *J. Guid., Control, Dyn.*, vol. 28, pp. 885–894, Sep./Oct. 2005.

[12] H. Sorenson and D. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, Jul. 1971.

[13] R. van der Merwe and E. Wan, "Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Hong Kong, Apr. 2003, pp. 701–704.

[14] J. Horwood and A. Poore, "Adaptive Gaussian sum filters for space surveillance," *IEEE Trans. Autom. Control*, vol. 56, no. 8, pp. 1777–1790, Aug. 2011.

[15] J. Horwood, N. Aragon, and A. Poore, "Gaussian sum filters for space surveillance: Theory and simulations," *J. Guid., Control, Dyn.*, vol. 34, pp. 1839–1851, Nov./Dec. 2011.

[16] G. Terejanu, P. Singla, T. Singh, and P. Scott, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE. Trans. Autom. Control*, vol. 56, no. 9, pp. 2151–2156, Sep. 2011.

[17] M. Psiaki, "The blind tricyclist problem and a comparative study of nonlinear filters," *IEEE Control Syst. Mag.*, vol. 33, no. 3, pp. 40–54, Jun. 2013.

[18] K. DeMars, R. Bishop, and M. Jah, "Entropy-based approach for uncertainty propagation of nonlinear dynamical systems," *J. Guid., Control, Dyn.*, vol. 36, pp. 1047–1057, Jul./Aug. 2013.

[19] E. Bayramoglu, O. Ravn, and N. Andersen, "A novel hypothesis splitting method implementation for multi-hypothesis filters," in *Proc. 10th IEEE Int. Conf. Control Autom.*, Hangzhou, China, Jun. 2013, pp. 574–579.

[20] M. Psiaki, J. Schoenberg, and I. Miller, "Gaussian sum reapproximation for use in a nonlinear filter," *J. Guid., Control, Dyn.*, vol. 38, pp. 292–303, Feb. 2015.

[21] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. New York, NY, USA: Academic, 1981.

[22] G. Bierman, *Factorization Methods for Discrete Sequential Estimation*. New York, NY, USA: Academic, 1977.

[23] M. Psiaki, "Blind tricyclist problem MATLAB functions, example input/output data file, and example use in EKF calculations," Dept. Mech. Aerosp. Eng., Cornell University, Ithaca, NY, USA, Jan. 2013. [Online]. Available: http://gps.mae.cornell.edu/blind_tricyclist_models_simulation_example.zip

Mark L. Psiaki received the B.A. degree in physics, the M.A. degree in mechanical and aerospace engineering, and the Ph.D. degree in mechanical and aerospace engineering, all from Princeton University, Princeton, NJ, USA. He joined the Department of Aerospace and Ocean Engineering at Virginia Tech, Blacksburg, VA, USA, in May 2016 and is currently a Professor and the Kevin Crofton Faculty Chair. He was a Faculty Member in the Sibley School of Mechanical and Aerospace Engineering, Cornell University, from 1986 to 2016, but he recently retired at the rank of Professor. He has conducted research in the areas of estimation and filtering, spacecraft attitude and orbit determination, GPS/GNSS receivers, navigation, and remote sensing using GNSS signals, GNSS security and integrity, aerospace vehicle guidance, numerical trajectory optimization, and dynamic modeling of spacecraft, aircraft, and wheeled vehicles. He recently completed a sabbatical leave as an NRC Senior Research Associate at the Air Force Research Lab, Kirtland AFB, NM, USA.