INVITED PAPER

Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature

New computationally efficient methods are proposed for more accurately analyzing and modeling dynamic processes that are nonlinear and subject to non-Gaussian noise.

By Ienkaran Arasaratnam, Simon Haykin, *Fellow IEEE*, and Robert J. Elliott

ABSTRACT | In this paper, a new version of the quadrature Kalman filter (QKF) is developed theoretically and tested experimentally. We first derive the new QKF for nonlinear systems with additive Gaussian noise by linearizing the process and measurement functions using statistical linear regression (SLR) through a set of Gauss–Hermite quadrature points that parameterize the Gaussian density. Moreover, we discuss how the new QKF can be extended and modified to take into account specific details of a given application. We then go on to extend the use of the new QKF to discrete-time, nonlinear systems with additive, possibly non-Gaussian noise. A bank of parallel QKFs, called the Gaussian sum-quadrature Kalman filter (GS-QKF) approximates the predicted and posterior densities as a finite number of weighted sums of Gaussian densities. The weights are obtained from the residuals of the QKFs. Three different Gaussian mixture reduction techniques are presented to alleviate the growing number of the Gaussian sum terms inherent to the GS-QKFs. Simulation results exhibit a significant improvement of the GS-QKFs over other nonlinear filtering approaches, namely, the basic bootstrap (particle) filters and Gaussian-sum extended Kalman filters, to solve nonlinear non-Gaussian filtering problems.

KEYWORDS | Gauss–Hermite Quadrature Rule; Gaussian sum filter; nonlinear filtering; quadrature Kalman filter; statistical linear regression (SLR)

Manuscript received October 2, 2006; revised February 7, 2007. This work is supported by the Natural Sciences & Engineering Research Council (NSERC) of Canada.
I. Arasaratnam and **S. Haykin** are with the Communications Research Laboratory, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: aienkaran@grads.ece.mcmaster.ca; haykin@mcmaster.ca).
R. J. Elliott is with the Haskayne School of Business Sciences, University of Calgary, Calgary, AL T2N 1N4, Canada (e-mail: relliott@ucalgary.ca).

Digital Object Identifier: 10.1109/JPROC.2007.894705

I. INTRODUCTION

The analysis and making of inferences about a dynamic system arise in a wide variety of applications in many disciplines. Examples include tracking the channel state information of a rapidly changing wireless channel [24], radar-based tracking of ships and aircraft [5], estimating the volatility of financial instruments using stock market data [1], and many others. The Bayesian framework is the most commonly used method for the study of these dynamic systems. In general, a Bayesian framework requires a dynamic state-space model (DSSM), which consists of two components: first, a process model describing the evolution of a hidden state of the system and second, a measurement model on noisy observables related to the hidden state. In the Bayesian approach, the posterior density of the state, obtained from Bayes' theorem, provides a complete statistical description of the state variable at that time [3].

A closed-form expression for the posterior density is available only for a restricted class of dynamic systems. For example, if the DSSM is linear with additive Gaussian noise and the prior distribution of the state variable is Gaussian, the predicted and posterior densities can be described by Gaussian densities. For this special case, the celebrated Kalman filter yields the optimal solution in the minimum-mean-square-error (MMSE) sense, the maximum likelihood (ML) sense, and the maximum *a posteriori* (MAP) sense [3], [18], [32]. However, the application of Bayesian estimation to real-world problems is plagued by two major difficulties: first, a realistic process and/or measurement model for a dynamic system of interest is often nonlinear, and second, the process noise and/or the measurement noise sources can be non-Gaussian.

In this paper, we develop approximate, recursive Gaussian filters for nonlinear dynamics with additive Gaussian or non-Gaussian noise:

1) A new quadrature Kalman filter (QKF) is developed for a nonlinear Gaussian dynamic system. Essentially, the new QKF is a rederivation of the existing QKF of Ito and Xiong [27] from a different perspective. This new QKF uses statistical linear regression (SLR) to linearize a nonlinear function through a set of Gauss–Hermite quadrature points. The new perspective allows us to make various extensions and modifications by taking into account specific details of a given application.
2) The new QKF algorithm is extended to a more general setting. A Gaussian-sum filter using a bank of new QKFs is employed for the general case of a nonlinear non-Gaussian dynamic system. The resulting algorithm is referred to as the Gaussian sum quadrature Kalman filter (GS-QKF). It is known that the non-Gaussian noise can be well approximated by a Gaussian-mixture[1] [2], [9], [24]. Hence, the non-Gaussian noise is modeled by a parallel combination of additive nonlinear Gaussian systems and consequently, the new QKF can be employed as a basic building block of GS-QKF.

The main issue in the GS-QKF design is that the number of terms in the Gaussian sum grows exponentially with time. In order to apply the GS-QKF in practice, we present three different Gaussian mixture reduction techniques to alleviate the growing memory problem.

The rest of the paper is organized as follows: Section II presents a general Bayesian framework for discrete-time nonlinear filtering. Section III highlights some of the key contributions in discrete-time nonlinear filtering since the development of the Kalman filter. Section IV briefly reviews the Gauss–Hermite quadrature rule, which is the basic building block of the new QKF. Section V presents the theory of SLR. In Section VI, the new QKF algorithm is developed using linear regression theory and the Gauss–Hermite quadrature rule. Section VII presents the GS-QKF filtering algorithm. Section VIII presents the Gaussian mixture reduction techniques. Section IX summarizes the GS-QKF filter. Section X analyzes the accuracy and the computational complexity of the Gauss–Hermite quadrature estimate. Section XI presents extended applications of the QKF. Section XII provides computer simulations, demonstrating the effectiveness of the new QKF algorithms over currently used methods for a nonlinear non-Gaussian filtering problem. Section XIII concludes the paper.

[1]The Gaussian mixture model has widely been used in the literature to model non-Gaussian noise sources. For example, in tracking a target in the presence of glint noise, the glint noise is modeled by a Gaussian mixture [9]. Similarly, in wireless communications, the measurement noise in an urban environment is modeled by a Gaussian mixture to capture the heavy tailed noise distribution due to outliers [24].

II. GENERAL BAYESIAN FRAMEWORK FOR NONLINEAR FILTERING

A continuous-time dynamic system, subject to random disturbances, can often be represented by the Itô stochastic differential equation [45]:

$$d\mathbf{x}_t = f(\mathbf{x}_t, \mathbf{u}_t, t)dt + g(\mathbf{x}_t, t)d\mathbf{v}_t \tag{1}$$

where $\mathbf{x}_t \in \mathbb{R}^{n_x}$ is a continuous-time signal, $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R} \rightarrow \mathbb{R}^{n_x}$ is a known function, $\mathbf{u}_t \in \mathbb{R}^{n_u}$ is the input vector, $\{\mathbf{v}_t, t \geq 0\}$ is an r-dimensional standard Brownian motion, and $g : \mathbb{R}^{n_x} \times \mathbb{R} \rightarrow \mathbb{R}^{n_x} \times \mathbb{R}^r$ is called the diffusion coefficient. The behavior of the system is observed imperfectly through the continuous observation process $\mathbf{z}_t \in \mathbb{R}^{n_z}$ that is related to the process $\mathbf{x}_t$ by the equation

$$\mathbf{z}_t = \int_0^t h(\mathbf{x}_s, \mathbf{u}_s, s)ds + \sqrt{R_t}\mathbf{w}_t \tag{2}$$

where $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R} \rightarrow \mathbb{R}^{n_z}$ is a known function, and $\{\mathbf{w}_t, t \geq 0\}$ is an n_z-dimensional standard Brownian motion, which is independent of $\mathbf{v}_t$ and the initial state $\mathbf{x}_0$. Write $\Im_t = \mathbb{B}(\mathbf{z}_s, s \leq t)$, where $\mathbb{B}(.)$ is the completion of the minimal σ-field over which the random variables in the parenthesis are measurable. The dynamics of the conditional density $p(\mathbf{x}_t|\Im_t)$ of $\mathbf{x}_t$, given $\Im_t$, are defined by the Kushner equation [36]:

$$dp(\mathbf{x}_t|\Im_t) = Lp(\mathbf{x}_t|\Im_t)dt + p(\mathbf{x}_t|\Im_t)R_t^{-1}(d\mathbf{z}_t - \hat{h}_t dt) \tag{3}$$

where

$$L = -\sum_{i=1}^{n_x} f_i(\mathbf{x}_t, \mathbf{u}_t, t)\frac{\partial}{\partial \mathbf{x}_i} + \frac{1}{2}\sum_{i,j=1}^{n_x} Q_{i,j}\frac{\partial^2}{\partial \mathbf{x}_i \partial \mathbf{x}_j}$$

$$\hat{h}_t = \mathbb{E}(h(\mathbf{x}_t, \mathbf{u}_t, t)|\Im_t)$$

$$Q_{i,j} = \{gg^T\}_{i,j}$$

We use the symbol $\mathbb{E}$ for the expectation operator throughout this paper.

A simpler solution to the above continuous-time filter is given by the Zakai equation [63]:

$$d\Pi(\mathbf{x}_t|\Im_t) = L\Pi(\mathbf{x}_t|\Im_t)dt + R_t^{-1}h(\mathbf{x}_t, \mathbf{u}_t, t)\Pi(\mathbf{x}_t|\Im_t)d\mathbf{z}_t \tag{4}$$

where $\Pi(\mathbf{x}_t|\Im_t)$ denotes an unnormalized conditional density of $\mathbf{x}_t$, given $\Im_t$. The Fokker–Planck operator L is defined as defined before. Although the Zakai equation and the Kushner equation have a one-to-one correspondence, the Zakai equation is simpler because it is linear in $\Pi(.)$. However, only for Gaussian densities that arise in the linear Gaussian case and for the Beneš type of nonlinearity, can the Zakai equation be solved explicitly [17], [25]. Numerous efforts have been devoted in the past to solve the Zakai equation for a general dynamic system. Approximate solutions to the Zakai equation can be found by applying numerical techniques. We cite the survey paper [22] and the references therein for the numerical techniques available in the literature. These methods are neither recursive nor computationally efficient.

When the continuous-time dynamic system is approximated by a discrete-time system, the derivatives in the continuous-time domain are approximated by difference equations in discrete-time. That is, consider

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, k) + \mathbf{v}_k \tag{5}$$

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{u}_k, k) + \mathbf{w}_k. \tag{6}$$

The nonlinear models given by (5) and (6) are respectively known as the process equation and the measurement (observation) equation. Using (5) and (6), the posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ at time k, where $\mathbf{z}_{1:j} = \{\mathbf{z}_1, \ldots, \mathbf{z}_j\}$, can be determined recursively in two steps: i) a time-update step and ii) a measurement update step.

Suppose the posterior density $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ at time k is known. Using (5), it is possible to obtain the predicted density of the state $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ at time k as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \tag{7}$$

which is called the time update.

On the receipt of a new measurement $\mathbf{z}_k$, the posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ can be written using Bayes' theorem as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{1}{c_k}p(\mathbf{x}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_k|\mathbf{x}_k). \tag{8}$$

Here, the normalizing constant c_k is expressed as

$$\begin{aligned} c_k &= p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) \\ &= \int_{\mathbb{R}^{n_x}} p(\mathbf{x}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_k|\mathbf{x}_k)d\mathbf{x}_k. \end{aligned} \tag{9}$$

Determination of the posterior density from the predicted density and the current measurement is called the measurement update. The filter is required to complete these two steps at every time step. The recurrence relationships (7) and (8) form the basis for the optimal Bayesian solution.[2] Knowledge of the posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ enables us to compute an optimal state estimate with respect to any criterion. For example, we may seek to find the optimal state estimate $\hat{\mathbf{x}}_{k|k}$ in the minimum-mean-squared-error (MMSE) sense.

$$\hat{\mathbf{x}}_{k|k} = \mathbb{E}(\mathbf{x}_k|\mathbf{z}_{1:k}) = \int_{\mathbb{R}^{n_x}} \mathbf{x}_k p(\mathbf{x}_k|\mathbf{z}_{1:k})d\mathbf{x}_k.$$

Similarly, the measure of accuracy of the state estimate, given by the covariance matrix $P_{k|k}$, is

$$\begin{aligned} P_{k|k} &= \mathbb{E}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T|\mathbf{z}_{1:k}\right] \\ &= \int_{\mathbb{R}^{n_x}} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T p(\mathbf{x}_k|\mathbf{z}_{1:k})d\mathbf{x}_k. \end{aligned}$$

However, for nonlinear filtering problems, the propagation of posterior density over time is only a conceptual solution, in the sense that in general it cannot be determined analytically. The implementation of the conceptual solution requires the storage of the entire probability density function. In general terms, this is equivalent to an infinite dimensional quantity except in a restricted set of cases, which can be exactly characterized by sufficient statistics of fixed and finite dimension. Since the analytical solution of (7) and (8) in most practical situations is intractable, we consider approximations to obtain a suboptimal solution.

[2]The recursive equations (7) and (8) can also be derived using the conditional expectations and change of measure (see [17] and [62] for more details). When estimating with Gaussian random variables, the conditional expectation is the projection onto the space of variables on which one conditions.

III. THE DEVELOPMENT OF NONLINEAR FILTERING: BRIEF HISTORY

The purpose of this section is to give a brief survey of major milestones in discrete-time nonlinear filtering, following the development of the Kalman filter. As described in Section I, the Kalman filter performs poorly when the assumptions of system linearity and/or Gaussianity of additive noise sources are violated [2], [52]. This has motivated intensive search for nonlinear filters for over four decades. As mentioned earlier, a computationally efficient, recursive optimal solution is available only for a limited class of dynamic systems. Subsequent investigations of nonlinear filtering have involved finding suboptimal solutions and may be classified into two major approaches: a local approach, approximating the posterior density function by some particular form, and a global approach, computing the posterior density function without making any explicit assumptions about its form.

A. Local Approach to Nonlinear Filtering

In the local approach to nonlinear filtering, we first mention an approximate Gaussian filter which is one of a broad class of suboptimal filters used in nonlinear, Gaussian filtering problems. Since the Gaussian filter approximates both the predicted density and the posterior density using Gaussian densities, it maintains the elegant recursive update form of the Kalman filter. An *extended Kalman filter* (EKF) is an example of such a filter [3], [5], [18], [28]. The EKF linearizes both the nonlinear process and the measurement dynamics with a first-order Taylor series expansion about the current state estimate, so that the Kalman filter recursions can be applied. Its accuracy, however, depends heavily on the "severity" of the nonlinearities; when the nonlinearities become severe, or they cannot be well approximated by a linear function, the EKF gives a divergent estimate [28], [30]. Since the Taylor series approximation of a nonlinear function is more accurate only in some neighborhood of a reference point, it fails to capture the global properties of that function. To improve the approximation, higher order terms of the Taylor series expansion can be retained. For example, a second-order EKF retains the Taylor series expansion up to a second term [5], [28]. However, it is questionable whether or not the higher order approximations improve the performance in cases where the standard EKF diverges. Furthermore, the second-order EKF requires Jacobians (first-order partial derivatives) and Hessians (second-order partial derivatives), whereas the standard EKF requires Jacobians only. Calculation of Jacobians and Hessians is often numerically unstable and computationally intensive. In some systems, the Jacobians and Hessians do not exist (e.g., for a process model representing abruptly changing behavior). Even if the function is differentiable, computing the Jacobian and Hessian may be hard if it is represented as a black-box rather than by some analytical form.

Another attempt to improve the performance of the EKF involves the use of an iterative update, and the resulting algorithm is called the *iterated Kalman filter* (IKF) [18], [28]. The basic idea of IKF is to linearize the measurement model around the updated state, rather than the predicted state. This is achieved iteratively, and it involves the use of the current measurement. It has been shown that the measurement update step of the IKF is more accurate only when the measurement model fully observes the state, and this is rarely the case in practice [39].

Recently, the *unscented* transformation has been used in the Kalman filter framework and the resulting filter is refereed to as the *unscented Kalman filter* [29]–[31]. The basic idea in the UKF is to choose deterministic sample (sigma), points that capture the mean and covariance of a Gaussian density. When propagated through a nonlinear function, these points capture the true mean and covariance up to a second-order of the nonlinear function. Instead of ignoring the high order terms, the UKF can also account for some of their effects by tuning a parameter used in the point selection. Despite its increased accuracy over the EKF, the sigma point scheme can be used with discontinuous functions as sigma points can straddle a discontinuity. For a state vector dimension higher than three, however, this scheme may require some "fine tuning" in order to prevent the propagation of a nonpositive definite covariance matrix. The *unscented* transformation is founded on intuition [30]:

> "It is easier to approximate a probability density than it is to approximate an arbitrary nonlinear function."

Another Gaussian filter, known as the *divided difference filter* (DDF) was introduced in [44] using Stirling's interpolations formula. Similar to the approach taken in the UKF, the DDF uses a deterministic sampling approach to propagate Gaussian statistics through a nonlinear function. Since the DDF shares a number of similarities with the UKF, such as deterministic sampling and weighted statistical estimation, both UKF and DDF algorithms are commonly referred to as sigma point filters [56].

Recently, Ito and Xiong introduced another suboptimal, nonlinear filter called the *quadrature Kalman filter* (QKF) [27]. The QKF uses the Gauss–Hermite numerical integration rule to calculate exactly the recursive Bayesian estimation integrals under the Gaussian assumption. Related works, presented in [14] and [35] on the use of Gauss–Hermite rule to approximate the moments of a conditional density in the context of nonlinear filtering, deserve recognition.

So far, we have discussed various algorithms based on Gaussian assumptions. In some nonlinear filtering cases, the densities can be multimodal or heavily skewed and can no longer be approximated by a single Gaussian. To

eliminate the Gaussian assumption, the Gram–Charlier or Edgeworth expansion can be used [53], [54]. This expansion is a series of Hermite polynomials which are orthogonal with respect to a Gaussian density, and it can be used to represent a wide class of density functions. This method seems to be more useful when the densities are unimodal though not Gaussian [52]. However, the main issue is that a large number of terms is required to obtain a reasonable approximation of a distinctly non-Gaussian density. Moreover, it is observed that the behavior of the estimator is sensitive to truncation of the infinite series [52].

Another approach taken in approximating the non-Gaussian density is a Gaussian-sum representation. Based on the fact that non-Gaussian densities can be approximated reasonably well by a finite sum of Gaussian densities (a Gaussian-sum), Alspach and Sorenson introduced the *Gaussian-sum filter* (GSF) for nonlinear systems [2]. In the Gaussian-sum approach, the key idea is to approximate both the predicted and the posterior densities as a sum of Gaussian densities, where the mean and covariance of each Gaussian density is calculated using the EKF algorithm. Essentially, the Gaussian sum filter of Alspach and Sorenson is a bank of EKFs running in parallel, where the state estimate is given by the weighted sum of the filters' outputs. The weights are estimated from the residuals of the EKFs. However, the main issue with the Gaussian sum approach is that the number of components in the Gaussian sum density grows exponentially with time.

B. Global Approach to Nonlinear Filtering

A second major direction taken in the literature on nonlinear filtering is a global approach that approximates the densities directly, so that the integrations involved in the Bayesian estimation framework are made as tractable as possible. The computational requirement of the global approach is, in general, greater than that of filters that assume a particular density such as the Gaussian density. However, the performance advantage offered by the global method may make the additional computational cost worthwhile.

The simplest method in the global approach is the point-mass method, where the densities are approximated by point masses located on a rectangular grid [10]. As a consequence, the recursive Bayesian estimation integrals can be evaluated numerically as a discrete, nonlinear convolution. A direct numerical computation of the mean and error covariance of the posterior density can also be evaluated by quadrature methods [14], [35], [59]. In the global approximation described in [33], piecewise constant functions are used to approximate the densities. This method is more sophisticated, and requires less computational cost than the point mass method. More sophisticated interpolation schemes, such as different spline approximations, have also been investigated in the literature [11], [15], [58].

Lastly, a relatively new technique called the *sequential Monte Carlo* (SMC) method uses a set of randomly chosen samples with associated weights to approximate the density [16], [21], [48]. Since the basic idea in the form of plain importance sampling degenerates over time, the SMC method includes a resampling step. As the number of samples becomes larger, the Monte Carlo characterization of the posterior density function becomes more accurate. However, the large number of samples often makes the use of SMC methods computationally expensive. Furthermore, the performance of SMC methods is crucially dependent on the selection of the so-called proposal distribution [16]. An intense research activity in the SMC field over more than a decade, however, has resulted in many improvements of the SMC methods and their applications. The SMC method consists of a large number of variants; many more are being constantly proposed in an attempt to improve both the statistical and computational efficiency (see [12], [13], [16], and the references therein).

The next section describes a local approach taken to approximate the integrals involved in the Bayesian filtering framework using the so-called Gauss–Hermite quadrature points. Consequently, a suboptimal solution to the state estimate is derived.

IV. GAUSS–HERMITE QUADRATURE RULE

Consider the following weighted integral of an integrable function $g(x)$ over the interval (a, b):

$$I(g) = \int_a^b W(x)g(x)dx$$

where $W(x)$ is a weight function which is positive almost everywhere except for a few points where it may be zero. An m-point numerical quadrature (integration) is an approximation of $I(g)$ of the form [47]:

$$I(g) \approx \sum_{l=1}^{m} \omega_l g(\xi_l).$$

Here, ξ_l are the quadrature points and ω_l are the associated weights. Given m distinct quadrature points, we can calculate ω_i by first computing the moments M_i of the integral

$$M_i = \int_a^b x^i W(x)dx, \quad \text{for } i \in \{0, 1, \ldots, (m-1)\},$$

and then solving the following Vandermonde system of equations [20]:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \xi_1 & \xi_2 & \dots & \xi_m \\ \vdots & \vdots & & \vdots \\ \xi_1^{m-1} & \xi_2^{m-1} & \dots & \xi_m^{m-1} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{pmatrix} = \begin{pmatrix} M_0 \\ M_1 \\ \vdots \\ M_{m-1} \end{pmatrix}. \quad (10)$$

The above nonlinear system of equations is also known as the system of moment equations. In the Gauss–Hermite quadrature rule, the weight function is chosen to be the standard Gaussian density with zero mean and unit variance $\mathcal{N}(x;0,1)$. The interval of interest is chosen to be $(-\infty,\infty)$. An obvious but algebraically difficult way of obtaining the quadrature points and weights is to seek solutions to the above nonlinear system. An alternative method is to take the quadrature points $\{\xi_i\}$ to be some appropriately chosen values and solve the above system for the weights $\{\omega_i\}$ only. According to the fundamental theorem of Gauss–Hermite quadrature [47], the quadrature points are chosen to be the zeros of the m-th order Hermite polynomial. Since the zeros of the Hermite polynomials are distinct, it is noteworthy that the determinant of the coefficient matrix in (10) is the well-known Vandermonde's determinant that is nonzero. Hence the solution vector $(\omega_1,\dots,\omega_m)$ is unique. Because we estimate $2m$ unknown parameters (m quadrature points and m associated weights), for an m-point quadrature scheme, the resulting quadrature rule is exact for all polynomials of degree $\leq (2m-1)$.

Consider a scalar random variable x having a Gaussian probability density $\mathcal{N}(x;0,1)$. The expected value of the function $g(x)$ can be approximated as

$$\begin{aligned} \mathbb{E}(g(x)) &= \int_{\mathbb{R}} g(x)\mathcal{N}(x;0,1)dx \\ &\approx \sum_{l=1}^{m} \omega_l g(\xi_l). \end{aligned} \quad (11)$$

Instead of finding quadrature points using root-finding methods, which may be mathematically unstable, a computationally better approach is presented in [19] to find the quadrature points and weights. This approach exploits the relationship between orthogonal polynomials and tridiagonals.

Suppose J is a symmetric tridiagonal matrix with zero diagonal elements and

$$J_{i,i+1} = \sqrt{i/2}, \quad 1 \leq i \leq (m-1).$$

Then the quadrature point ξ_l is taken to be $\xi_l = \sqrt{2}\varepsilon_l$, where ε_l is the l-th eigenvalue of J; and the corresponding weight $\omega_l = (\nu_l)_1^2$, where $(\nu_l)_1$ is the first element of the l-th normalized eigenvector of J.

Now consider a vector-valued random variable $\mathbf{x}$ having a Gaussian density $p(\mathbf{x}) = \mathcal{N}(\mathbf{x};\mathbf{0},I_{n_x})$, where I_{n_x} is the identity matrix of dimensions $n_x \times n_x$. Since the individual components of $\mathbf{x}$ are mutually uncorrelated, the 1-D quadrature formula in (11) can be extended to the multidimensional quadrature formula by successively applying it to compute the expectation

$$\begin{aligned} \mathbb{E}(g(\mathbf{x})) &= \int_{\mathbb{R}^{n_x}} g(\mathbf{x})\mathcal{N}(\mathbf{x};\mathbf{0},I_{n_x})d\mathbf{x} \\ &\approx \sum_{l_{n_x}=1}^{m} \omega_{l_{n_x}} \dots \sum_{l_1=1}^{m} \omega_{l_1} g\left(\xi_{l_1} \dots \xi_{l_{n_x}}\right) \\ &= \sum_{l=1}^{m^{n_x}} \omega_l g(\xi_l). \end{aligned} \quad (12)$$

Here $\xi_l = [\xi_{l_1} \dots \xi_{l_{n_x}}]^T$ and $\omega_l = \prod_{j=1}^{n_x} \omega_{l_j}$.

On the other hand, if we consider the probability density function $p(\mathbf{x}) = \mathcal{N}(\mathbf{x};\bar{\mathbf{x}},\Sigma)$, the individual components are mutually correlated and the cartesian product rule of (12) cannot be efficiently applied [51]. This issue is illustrated in Fig. 1 for the 2-D case. The stylized contours in Fig. 1(a) indicate that the lattice of integration points formed from the product of two 1-D grids does not efficiently cover the bulk of the density function. This is due to the high correlation among individual components of the density, and the correlation leads to many of the lattice points falling in the areas of negligible density, thus yielding a poor estimate of the integration. The density in Fig. 1(b) is obtained from the density in Fig. 1(a) when the *stochastic decoupling* [50] technique is applied. This technique essentially eliminates the correlation of the original density, and thus improves the spread of the density equally in all directions. We now describe how the stochastic decoupling technique is exploited to improve the estimate of an integration.

Suppose that the square-root[3] of the covariance matrix, Σ is written as $\sqrt{\Sigma}$

$$\Sigma = \sqrt{\Sigma}(\sqrt{\Sigma})^T.$$

[3]There are many types of matrix decomposition techniques that factorize a covariance matrix P in the form of $P = SS^T$ [8]. E.g., the Cholesky decomposition, the singular value decomposition (SVD) and the eigenvector decomposition. Choosing which one to use primarily depends upon the particular application, numerical concerns and desired level of accuracy. For example, the SVD is the most robust algorithm to factorize a covariance matrix especially when the covariance matrix P becomes nearly singular [8]. However, it is also one of the most computationally expensive algorithms. In the case of SVD, the square-root factor of P, S can be chosen to be $S = U\sqrt{D}$, where the SVD of P can be expressed as $P = UDU^T$ with D a diagonal matrix.

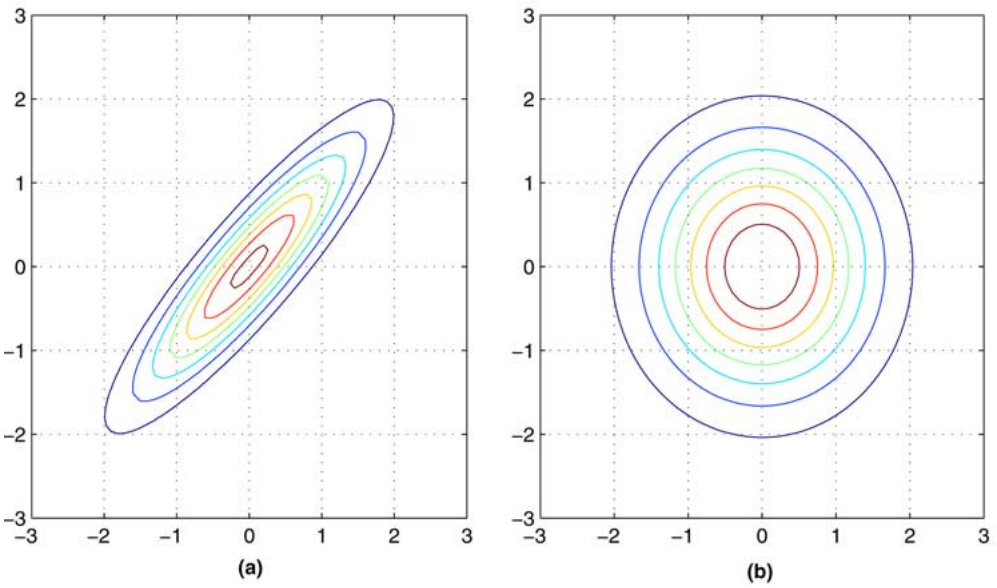

Fig. 1. ***Contour plots of zero mean Gaussian densities with covariances [1 0.9; 0.9 1] [see Fig. 1(a)] and [1 0; 0 1] [see Fig. 1(b)] respectively.***

where the superscript T denotes the matrix transpose. By applying the linear transformation (coordinate rotation) and writing

$$\mathbf{y} = (\sqrt{\Sigma})^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \tag{13}$$

the stochastic decoupling technique yields a new density with uncorrelated elements of unity variance as shown in Fig. 1(b). Here A^{-1} refers to the inverse of matrix A. Now, the Cartesian product rule can be efficiently used to evaluate the expectation:

$$\begin{aligned}\mathbb{E}(g(\mathbf{x})) &= \int_{\mathbb{R}^{n_x}} g(\mathbf{x})\mathcal{N}(\mathbf{x};\bar{\mathbf{x}},\Sigma)d\mathbf{x} \\ &= \int_{\mathbb{R}^{n_x}} g(x)\frac{1}{\sqrt{|2\pi\Sigma_{\mathbf{x}}|}} \\ &\quad \times \exp\left[-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\Sigma_{\mathbf{x}}^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right]d\mathbf{x}. \end{aligned} \tag{14}$$

Now the Jacobian of transformation $\nabla_x\mathbf{y}$ equals $(\sqrt{\Sigma})^{-1}$, so substituting (13) into (14) yields

$$\begin{aligned}\mathbb{E}(g(\mathbf{x})) &= \int_{\mathbb{R}^{n_x}} g(\sqrt{\Sigma}\mathbf{y}+\bar{\mathbf{x}})\frac{1}{(2\pi)^{\frac{n_x}{2}}}\exp\left[-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right]d\mathbf{y} \\ &= \int_{\mathbb{R}^{n_x}} g(\sqrt{\Sigma}\mathbf{y}+\bar{\mathbf{x}})\mathcal{N}(\mathbf{y};\mathbf{0},I_{n_x})d\mathbf{y}. \end{aligned} \tag{15}$$

Since (15) takes a form similar to (12), we can now apply the Gauss–Hermite quadrature rule to approximate (15) as follows:

$$\begin{aligned}\mathbb{E}(g(\mathbf{x})) &= \sum_{l_{n_x}=1}^{m}\omega_{l_{n_x}}\cdots\sum_{l_1=1}^{m}\omega_{l_1}g\left(\sqrt{\Sigma}\left[\xi_{l_1}\ldots\xi_{l_{n_x}}\right]^T+\bar{\mathbf{x}}\right) \\ &= \sum_{l=1}^{m^{n_x}}\omega_l\, g(\sqrt{\Sigma}\xi_l+\bar{\mathbf{x}}). \end{aligned} \tag{16}$$

Let $\chi_l = \sqrt{\Sigma}\xi_l + \bar{\mathbf{x}}$. We can therefore rewrite $\mathbb{E}(g(\mathbf{x}))$ in (16) as

$$\mathbb{E}(g(\mathbf{x})) = \sum_{l=1}^{m^{n_x}}\omega_l g(\chi_l). \tag{17}$$

As $\sum_{l=1}^{m^{n_x}}\omega_l = 1$ and $\omega_l > 0$ for $l = 1, 2, \ldots, m^{n_x}$, the right side of (17) resembles the expectation of a function of the random variable χ such that

$$\mathbb{E}(g(\mathbf{x})) = \sum_{l=1}^{m^{n_x}}\omega_l g(\chi_l) = \sum_{l=1}^{m^{n_x}} g(\chi_l)p\{\chi=\chi_l\},$$

where the probability $p\{\chi = \chi_l\} = \omega_l$. This insight and the SLR, described later in Section V, are applied to develop the new QKF algorithm in Section VI.

V. STATISTICAL LINEAR REGRESSION

This section briefly reviews the background of SLR theory, which is the basic building block of our new filtering algorithm presented in Section VI. The SLR takes into account the uncertainty, or "probabilistic spread," of the prior random variable when linearizing a nonlinear function, which operates on that random variable. By doing so, the resulting linearized function is more accurate in a statistical sense than simply using a first-order truncated Taylor series expansion of the function around a single point, say a prior mean [18].

Suppose $\mathbf{x}$ is an $n_{\mathbf{x}}$ dimensional random variable with a Gaussian density having mean $\bar{\mathbf{x}}$ and covariance P_{xx}. Suppose a second random variable $\mathbf{y}$ is related to $\mathbf{x}$ through the nonlinear function

$$\mathbf{y} = g(\mathbf{x}).$$

The objective is to obtain a linear estimator of $\mathbf{y}$, $\hat{\mathbf{y}}$ such that

$$\hat{\mathbf{y}} = A\mathbf{x} + \mathbf{b}. \tag{18}$$

Here A and $\mathbf{b}$ are respectively a matrix and vector which are to be determined by minimizing the mean-squared error as the criterion function

$$\{A, \mathbf{b}\} = \arg\min \mathbb{E}(\mathbf{e}^T\mathbf{e}). \tag{19}$$

where $\mathbf{e}$ is the linearization error defined as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}. \tag{20}$$

The following exposition follows closely that of [18] (chap. 6, pp. 203–209) and [38]. Substituting (20) into (19) and setting the partial derivative of (19) with respect to $\mathbf{b}$ to zero, we obtain

$$(-2)\mathbb{E}(\mathbf{y} - A\mathbf{x} - \mathbf{b}) = \mathbf{0}. \tag{21}$$

Solving (21) for $\mathbf{b}$ gives

$$\mathbf{b} = \bar{\mathbf{y}} - A\bar{\mathbf{x}}. \tag{22}$$

Here $\bar{\mathbf{x}} = \mathbb{E}(\mathbf{x})$ and $\bar{\mathbf{y}} = \mathbb{E}(\mathbf{y})$. Substituting $\mathbf{b}$ from (22) into (19) and setting the gradient with respect to A to zero, we obtain

$$(-2)\mathbb{E}\big([(\mathbf{y} - \bar{\mathbf{y}}) - A(\mathbf{x} - \bar{\mathbf{x}})][\mathbf{x} - \bar{\mathbf{x}}]^T\big) = \mathbf{0}. \tag{23}$$

Solving (23) for A yields

$$A = P_{xy}^T P_{xx}^{-1} \tag{24}$$

where $P_{xx} = \mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$ and $P_{xy} = \mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^T]$. The equation for the regression line is therefore given by

$$\hat{\mathbf{y}} = A\mathbf{x} + \mathbf{b} \tag{25}$$

where the regression coefficients A and $\mathbf{b}$ are given by (24) and (22). The mean error $\bar{\mathbf{e}}$ is given by

$$\begin{aligned}\bar{\mathbf{e}} = \mathbb{E}(\mathbf{e}) &= \mathbb{E}(\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbb{E}[\mathbf{y} - (A\mathbf{x} + \mathbf{b})] \\ &= \bar{\mathbf{y}} - A\bar{\mathbf{x}} - \mathbf{b} \\ &= \mathbf{0}.\end{aligned} \tag{26}$$

The error covariance matrix P_{ee} is given by

$$\begin{aligned}P_{ee} &= \mathbb{E}(\mathbf{e}\mathbf{e}^T) \\ &= \mathbb{E}[(\mathbf{y} - \bar{\mathbf{y}}) - A(\mathbf{x} - \bar{\mathbf{x}})][(\mathbf{y} - \bar{\mathbf{y}}) - A(\mathbf{x} - \bar{\mathbf{x}})]^T \\ &= P_{yy} - AP_{xy} - P_{yx}A^T + AP_{xx}A^T \\ &= P_{yy} - AP_{xx}A^T.\end{aligned} \tag{27}$$

It is noteworthy here that the regression coefficients can be derived in another way using the *principle of orthogonality* [34].

A. Evaluation of Regression Coefficients Using Gauss–Hermite Quadrature Points

For a Gaussian random variable $\mathbf{x}$ with mean $\bar{\mathbf{x}}$ and covariance P_{xx}, we can choose the 3-point (per axis) Gauss–Hermite quadrature point set, which obeys the following conditions:

$$\bar{\mathbf{x}} = \mathbb{E}(\mathbf{x}) = \sum_{l=1}^{m} \omega_l X_l \tag{28}$$

$$\begin{aligned}P_{xx} &= \mathbb{E}\big[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\big] \\ &= \sum_{l=1}^{m} \omega_l (X_l - \bar{\mathbf{x}})(X_l - \bar{\mathbf{x}})^T.\end{aligned} \tag{29}$$

For the 3-point Gauss–Hermite quadrature, m is given by $m = 3^{n_x}$. Moreover, as described in Section IV, the statistics of the random variable $\mathbf{y}$, which is related to $\mathbf{x}$ by $\mathbf{y} = g(\mathbf{x})$, can be approximately calculated from $\{\omega_l, X_l\}_{l=1}^{m}$ using the Gauss–Hermite quadrature rule

$$\bar{\mathbf{y}} = \mathbb{E}(\mathbf{y}) \approx \sum_{l=1}^{m} \omega_l Y_l \tag{30}$$

$$P_{yy} = \mathbb{E}\left[(\mathbf{y}-\mathbf{y})(\mathbf{y}-\mathbf{y})^T\right] \approx \sum_{l=1}^{m} \omega_l (Y_l - \bar{\mathbf{y}})(Y_l - \bar{\mathbf{y}})^T \tag{31}$$

$$P_{xy} = \mathbb{E}\left[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{y}-\bar{\mathbf{y}})^T\right] \approx \sum_{l=1}^{m} \omega_l (X_l - \bar{\mathbf{x}})(Y_l - \bar{\mathbf{y}})^T. \tag{32}$$

Here $Y_l = g(X_l)$. Given the set of weighted points $\{X_l, Y_l\}_{l=1}^{m}$ and the associated weights or probabilities $\{\omega_l\}_{l=1}^{m}$, we can find the linear regression of $\mathbf{y}$ on $\mathbf{x}$ given by

$$\hat{\mathbf{y}} = A\mathbf{x} + \mathbf{b} \tag{33}$$

where

$$A = P_{xy}^T P_{xx}^{-1} = \left[\sum_{l=1}^{m} \omega_l (Y_l - \bar{\mathbf{y}})(X_l - \bar{\mathbf{x}})^T\right] \times \left[\sum_{l=1}^{m} \omega_l (X_l - \bar{\mathbf{x}})(X_l - \bar{\mathbf{x}})^T\right]^{-1}$$

$$\mathbf{b} = \bar{\mathbf{y}} - A\bar{\mathbf{x}}.$$

Here $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are given by (28) and (30). Since the estimate $\hat{\mathbf{y}}$ is a random variable, its mean is given by

$$\mathbb{E}(\hat{\mathbf{y}}) = A\bar{\mathbf{x}} + \mathbf{b} = \bar{\mathbf{y}} = \sum_{l=1}^{m} \omega_l Y_l. \tag{34}$$

The error covariance is given by

$$\operatorname{cov}(\hat{\mathbf{y}}) = AP_{xx}A^T = P_{yy} - P_{ee} = \sum_{l=1}^{m} \omega_l (Y_l - \bar{\mathbf{y}})(Y_l - \bar{\mathbf{y}})^T - P_{ee}. \tag{35}$$

In summary, this section shows how the statistics computed via the Gauss–Hermite quadrature rule can be used to linearize a nonlinear function $g(\mathbf{x})$ in the minimum-mean-squared-error sense, when a Gaussian random variable $\mathbf{x}$ is provided as its argument. The regression coefficients A and $\mathbf{b}$ of the linearized function of $g(\mathbf{x})$, $A\mathbf{x} + \mathbf{b}$, are determined from the Gauss–Hermite quadrature points. In the next section, the linearization technique is applied to nonlinear process and measurement functions; thus a new QKF algorithm in the Kalman filter framework is derived.

Remark:

- From (27), we have $P_{yy} = P_{ee} + AP_{xx}A^T$. This implies that the more severe the nonlinearity is over the "uncertainty region" of $\mathbf{x}$, the larger the error covariance of $\mathbf{y}$, P_{yy}, will be and accordingly the estimate of $\mathbf{y}$, $\hat{\mathbf{y}}$, will be less accurate. Moreover, it is noteworthy that the error covariance $P_{yy} = \sum_{l=1}^{m} \omega_l (Y_l - \bar{\mathbf{y}})(Y_l - \bar{\mathbf{y}})^T$ [from (31)], indirectly incorporates the effect of the linearization error because of (27). Consequently, the SLR technique is proved to yield a consistent estimate. The estimate $\hat{\mathbf{y}}$ is said to be consistent if $P_{yy} \geq \mathbb{E}(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T)$, where $\tilde{\mathbf{y}}$ is the true error in the estimate $\hat{\mathbf{y}}$.

VI. NEW QUADRATURE KALMAN FILTERING ALGORITHM

This section presents a systematic development of the new QKF algorithm from a linear regression perspective. Here, the regression points are Gauss–Hermite quadrature points that can be used to parameterize a Gaussian density. We assume that the process noise sequence $\mathbf{v}_k$ and the measurement noise sequence $\mathbf{w}_k$ are both Gaussian with zero mean, and

$$\mathbb{E}\left(\mathbf{v}_k \mathbf{v}_j^T\right) = Q_k \delta_{kj}$$
$$\mathbb{E}\left(\mathbf{w}_k \mathbf{w}_j^T\right) = R_k \delta_{kj}$$
$$\mathbb{E}\left(\mathbf{v}_k \mathbf{w}_j^T\right) = 0 \quad \text{for all } k, j.$$

Here δ is the Kronecker delta function. The initial state vector $\mathbf{x}_0$ is assumed to be described by a known density

function $p(\mathbf{x}_0)$. We derive the new QKF by linearizing the process and measurement functions by a linear regression using the set of Gauss–Hermite quadrature points. As a result, we obtain a closed-form solution for the predicted and posterior densities. The remainder of the section presents the new QKF algorithm consisting of two steps: i) the time update and ii) the measurement update.

A. Time Update of the New QKF

We consider (5) without the additive noise component for the moment and incorporate its effect into the new QKF algorithm at the end of this subsection. The new QKF algorithm first finds the m-regression, or quadrature points $\{X_{l,k-1|k-1}\}_{l=1}^{m}$ with the associated weights $\{\omega_l\}_{l=1}^{m}$ in the state space, so that these points have the mean $\hat{\mathbf{x}}_{k-1|k-1}$ and covariance $P_{k-1|k-1}$.

Write:

$$X_{l,k-1|k-1} = \sqrt{P_{k-1|k-1}}\xi_l + \hat{\mathbf{x}}_{k-1|k-1}. \tag{36}$$

Here $\{\xi_l\}$ are Gauss–Hermite quadrature points with weights $\{\omega_l\}$, and they approximate the standard Gaussian density. The one-step predicted regression points $\{X^*_{l,k|k-1}\}$ are obtained by propagating $\{X_{l,k-1|k-1}\}$ through the process function $f(.)$:

$$\begin{aligned} X^*_{l,k|k-1} &= f(X_{l,k-1|k-1}, \mathbf{u}_{k-1}, k-1) \\ &\text{for } l = 1, 2, \ldots, m. \end{aligned} \tag{37}$$

Since the new QKF is developed in the Kalman filter framework, it uses a linearized process function obtained by the SLR through $\{X_{l,k-1|k-1}, X^*_{l,k|k-1}\}_{l=1}^{m}$. That is, (5) can now be expressed by the following linear function:

$$\begin{aligned} \mathbf{x}_k &= f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, k-1) \\ &= \underbrace{A_{f,k-1}\mathbf{x}_{k-1} + \mathbf{b}_{f,k-1}}_{\text{LA}} + \underbrace{\mathbf{e}_k}_{\text{LE}}. \end{aligned} \tag{38}$$

Here "LA" and "LE" stand for the linear approximation and linearization error, respectively. As described in the previous section, $A_{f,k-1}$ and $\mathbf{b}_{f,k-1}$ are, respectively, given by

$$A_{f,k-1} = P^T_{x_{k-1}x_k} P^{-1}_{k-1|k-1} \tag{39}$$

$$\text{and} \quad \mathbf{b}_{f,k-1} = \hat{\mathbf{x}}_{k|k-1} - A_{f,k-1}\hat{\mathbf{x}}_{k-1|k-1}. \tag{40}$$

Here

$$\begin{aligned} P_{x_{k-1}x_k} &= \mathbb{E}\left[(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T\right] \\ &= \sum_{l=1}^{m} \omega_l (X_{l,k-1|k-1} - \hat{\mathbf{x}}_{k-1|k-1})\left(X^*_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right)^T. \end{aligned}$$

By taking the expectation of (38) conditioned on the measurements obtained up to $(k-1)$, the predicted state estimate $\hat{\mathbf{x}}_{k|k-1}$ is computed as

$$\hat{\mathbf{x}}_{k|k-1} = A_{f,k-1}\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{b}_{f,k-1} + \mathbf{0}. \tag{41}$$

$\hat{\mathbf{x}}_{k|k-1}$ in (41) can be written as the weighted sum of regression points [compare (34) with (41)]

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{l=1}^{m} \omega_l X^*_{l,k|k-1}. \tag{42}$$

From (38), the predicted error covariance $P_{k|k-1}$ is computed as

$$P_{k|k-1} = A^T_{f,k-1} P_{k-1|k-1} A^T_{f,k-1} + P_{ee}. \tag{43}$$

$P_{k|k-1}$ in (43) can be written as the weighted sum of regression points [compare (35) with (43)]

$$P_{k|k-1} = \sum_{l=1}^{m} \omega_l \left(X^*_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right)\left(X^*_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right)^T. \tag{44}$$

Since the sum of quadrature weights equals unity, the expression for the predicted error covariance in (44) can be simplified further as

$$P_{k|k-1} = \sum_{l=1}^{m} \omega_l X^*_{l,k|k-1} X^{*T}_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{x}}^T_{k|k-1}. \tag{45}$$

For an additive Gaussian process noise with zero mean and covariance Q_k, the expression for the predicted mean is given by (41), while the predicted error covariance in (45) is expanded by adding the term Q_k, as shown by

$$P_{k|k-1} = Q_k + \sum_{l=1}^{m} \omega_l X^*_{l,k|k-1} X^{*T}_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{x}}^T_{k|k-1}. \tag{46}$$

Therefore, at the end of the time-update step, the new QKF approximates the predicted density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, given by $p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})p(\mathbf{x}_k|\mathbf{x}_{k-1})d\mathbf{x}_{k-1}$, as a Gaussian density with mean $\hat{\mathbf{x}}_{k|k-1}$ and covariance $P_{k|k-1}$.

B. Measurement Update of the New QKF

Since the points $\{X^*_{l,k|k-1}, \omega_l\}_{l=1}^m$ do not reflect the added uncertainty due to process noise, the new QKF finds m-regression or quadrature points $\{X_{l,k|k-1}\}_{l=1}^m$ with the associated weights $\{\omega_l\}_{l=1}^m$ that have the mean $\hat{\mathbf{x}}_{k|k-1}$ and covariance $P_{k|k-1}$

$$X_{l,k|k-1} = \sqrt{P_{k|k-1}}\xi_l + \hat{\mathbf{x}}_{k|k-1} \quad l = 1 \ldots m. \tag{47}$$

We then linearize the measurement function $h(.)$ in a manner similar to the process function as in Section VI-A. We summarize the measurement update steps of the new QKF as follows.

i) Starting from $\{X_{l,k|k-1}\}_{l=1}^m$, the one-step predicted regression points are propagated through the measurement model (6) as

$$Z_{l,k|k-1} = h(X_{l,k|k-1}, \mathbf{u}_k, k), \quad l = 1, 2, \ldots m. \tag{48}$$

ii) The new QKF uses a linearized measurement model obtained by the linear regression through $\{X_{l,k|k-1}, Z_{l,k|k-1}\}_{l=1}^m$. The estimate of the predicted measurement $\hat{\mathbf{z}}_{k|k-1}$ is computed as a weighted sum of regression points

$$\hat{\mathbf{z}}_{k|k-1} = \sum_{l=1}^m \omega_l Z_{l,k|k-1}. \tag{49}$$

iii) For an additive measurement noise covariance R_k, the innovation covariance is computed as a weighted sum of outer products of regression points

$$\begin{aligned} P_{zz,k|k-1} &= R_k + \sum_{l=1}^m \omega_l (Z_{l,k|k-1} - \hat{\mathbf{z}}_{k|k-1}) \\ &\quad \times (Z_{l,k|k-1} - \hat{\mathbf{z}}_{k|k-1})^T \\ &= R_k + \sum_{l=1}^m \omega_l Z_{l,k|k-1} Z^T_{l,k|k-1} \\ &\quad - \hat{\mathbf{z}}_{k|k-1}\hat{\mathbf{z}}^T_{k|k-1}. \end{aligned} \tag{50}$$

iv) The cross-covariance matrix is computed as

$$\begin{aligned} P_{xz,k|k-1} &= \sum_{l=1}^m \omega_l (X_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1})(Z_{l,k|k-1} - \hat{\mathbf{z}}_{k|k-1})^T \\ &= \sum_{l=1}^m \omega_l X_{l,k|k-1} Z^T_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{z}}^T_{k|k-1}. \end{aligned} \tag{51}$$

Since $[\mathbf{x}_k \ \mathbf{z}_k]^T$ is assumed to be jointly Gaussian, its sufficient statistics can be expressed by the mean $[\hat{\mathbf{x}}_{k|k-1} \ \hat{\mathbf{z}}_{k|k-1}]^T$ and covariance $\begin{pmatrix} P_{k|k-1} & P_{xz,k|k-1} \\ P_{zx,k|k-1} & P_{zz,k|k-1} \end{pmatrix}$ respectively. Consequently, the posterior (conditional) density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ at time k can be computed recursively as [3]

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, P_{k|k}). \tag{52}$$

Here, we have

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + W_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \tag{53}$$

and

$$P_{k|k} = P_{k|k-1} - W_k P_{zz,k|k-1} W_k^T. \tag{54}$$

The Kalman gain W_k is defined as

$$W_k = P_{xz,k|k-1} P^{-1}_{zz,k|k-1}.$$

C. Summary of the New QKF Algorithm

This subsection summarizes the new QKF algorithm that computes both the time update and measurement update steps at each time-step. Though the new QKF is algebraically equivalent to the QKF of Ito and Xiong, it is in fact a more simplified version as it views the nonlinear filtering problem from the SLR perspective rather than the numerical integration perspective.

Time Update Step

1) Assume at time k that the posterior density function $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k-1|k-1}, P_{k-1|k-1})$ is known. Factorize

$$P_{k-1|k-1} = \sqrt{P_{k-1|k-1}}(\sqrt{P_{k-1|k-1}})^T.$$

2) Evaluate the quadrature points $\{X_{l,k-1|k-1}\}_{l=1}^{m}$ as:

$$X_{l,k-1|k-1} = \sqrt{P_{k-1|k-1}}\xi_l + \hat{\mathbf{x}}_{k-1|k-1}.$$

3) Evaluate the propagated quadrature points $\{X^*_{l,k|k-1}\}_{l=1}^{m}$ as:

$$X^*_{l,k|k-1} = f(X_{l,k-1|k-1}, \mathbf{u}_{k-1}, k-1).$$

4) Estimate the predicted state:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{l=1}^{m} \omega_l X^*_{l,k|k-1}.$$

5) Estimate the predicted error covariance:

$$P_{k|k-1} = \sum_{l=1}^{m} \omega_l X^*_{l,k|k-1} X^{*T}_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{x}}^T_{k|k-1} + Q_k.$$

At the end of the time update, we have the predicted density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k-1}, P_{k|k-1})$.

Measurement Update Step

1) Factorize

$$P_{k|k-1} = \sqrt{P_{k|k-1}}\left(\sqrt{P_{k|k-1}}\right)^T.$$

2) Evaluate the quadrature points $\{X_{l,k|k-1}\}_{l=1}^{m}$ as:

$$X_{l,k|k-1} = \sqrt{P_{k|k-1}}\xi_l + \hat{\mathbf{x}}_{k|k-1}.$$

3) Evaluate the propagated quadrature points $\{Z_{l,k|k-1}\}_{l=1}^{m}$ as:

$$Z_{l,k|k-1} = h(X_{l,k|k-1}, \mathbf{u}_k, k).$$

4) Estimate the predicted measurement:

$$\hat{\mathbf{z}}_{k|k-1} = \sum_{l=1}^{m} \omega_l Z_{l,k|k-1}.$$

5) Estimate the innovation covariance matrix:

$$P_{zz,k|k-1} = R_k + \sum_{l=1}^{m} \omega_l Z_{l,k|k-1} Z^T_{l,k|k-1} - \hat{\mathbf{z}}_{k|k-1}\hat{\mathbf{z}}^T_{k|k-1}.$$

6) Estimate the cross covariance matrix:

$$P_{xz,k|k-1} = \sum_{l=1}^{m} \omega_l X_{l,k|k-1} Z^T_{l,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{z}}^T_{k|k-1}.$$

7) Estimate the Kalman gain:

$$W_k = P_{xz,k|k-1} P^{-1}_{zz,k|k-1}.$$

8) Estimate the updated state:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + W_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}).$$

9) Estimate the corresponding error covariance:

$$P_{k|k} = P_{k|k-1} - W_k P_{zz,k|k-1} W^T_k.$$

At the end of the measurement update, we have the posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, P_{k|k})$.

VII. GAUSSIAN SUM-QUADRATURE KALMAN FILTER ALGORITHM

So far we have focussed on nonlinear filtering problems with additive Gaussian noise. However, in practice, difficulties with the more general setting arise from two sources. The first is the occurrence of non-Gaussian process and/or measurement noise and the second is non-Gaussian prior information. In this section, our investigations are directed towards developing methods of Gaussian sum approximations that allow approximate solution of nonlinear filtering problems. Consequently, the QKF can be applied for computing the predicted and posterior densities as Gaussian sums. By doing so, we derive the algorithm called Gaussian sum-quadrature Kalman filter (GS-QKF).

A. Modeling of Non-Gaussian Density

Several models have been used to date to model non-Gaussian noise environments. Some of these models have

been developed directly from the underlying physical phenomenon, most notably the Middleton class A, B, and C models [41]. On the other hand, empirically devised noise models have been used over the years to approximate many non-Gaussian noise distributions. Based on the Wiener approximation theorem, any non-Gaussian noise distribution can be expressed as, or approximated sufficiently well by, a finite sum of known Gaussian densities.

The so-called Gaussian sum approach is summarized in the following lemma [3]:

Lemma: Any density $p(\mathbf{x})$ associated with an n-dimensional vector $\mathbf{x}$ can be approximated as closely as desired by a density of the form

$$p_A(\mathbf{x}) = \sum_{i=1}^{n} a_i \mathcal{N}(\bar{\mathbf{x}}_i, \Sigma_i)$$

for some integer n, and positive scalars a_i with $\sum_{i=1}^{n} a_i = 1$.

It can be shown that the density $p_A(\mathbf{x})$ converges uniformly to any density function of practical interest by letting the number of terms increase and each elemental covariance approach to zero [3].

This approximation procedure has been used to develop empirical densities which relate to many physical non-Gaussian phenomena. The ε-mixture noise model has been extensively used to describe a non-Gaussian noise environments in communication and control systems, target tracking in the presence of glint noise, jamming or clutter suppression, outlier rejection in image processing applications, and intelligent processing in interferometric and multirage measurement systems.

Non-Gaussian noise densities can be approximated empirically by Gaussian-sums as closely as possible using Gaussian mixture learning algorithms such as the expectation-maximization (EM) algorithm and the k-means algorithm [57].[4] Consequently, for a state-space model, such as the one described in (5) and (6) with a Gaussian-sum additive noise sources, it is possible to obtain both the predicted and posterior densities as Gaussian-sums. Assume at time k that the additive process and measurement noise are both available as approximate Gaussian sums

$$\begin{aligned} p(\mathbf{v}_k) &\approx \sum_{i=1}^{l_1} \beta_{ki} p_i(\mathbf{v}_k) \\ &= \sum_{i=1}^{l_1} \beta_{ki} \mathcal{N}(\mathbf{v}_k; \bar{\mathbf{v}}_{ki}, Q_{ki}) \end{aligned} \tag{55}$$

$$\begin{aligned} p(\mathbf{w}_k) &\approx \sum_{i=1}^{l_2} \mu_{ki} p_i(\mathbf{w}_k) \\ &= \sum_{i=1}^{l_2} \mu_{ki} \mathcal{N}(\mathbf{w}_k; \bar{\mathbf{w}}_{ki}, R_{ki}) \end{aligned} \tag{56}$$

where, β_{ki} and μ_{ki} are nonnegative constants satisfying

$$\sum_{i=1}^{l_1} \beta_{ki} = 1 \text{ and } \sum_{i=1}^{l_2} \mu_{ki} = 1.$$

Moreover, the prior density at time zero is also assumed to be a Gaussian sum

$$\begin{aligned} p(\mathbf{x}_0) &= \sum_{i=1}^{l_0} \alpha_{0i} p_i(\mathbf{x}_0) \\ &= \sum_{i=1}^{l_0} \alpha_{0i} \mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_{0i}, \bar{P}_0). \end{aligned} \tag{57}$$

Here, α_{0i} are nonnegative constants and $\sum_{i=1}^{l_0} \alpha_{0i} = 1$. The GS-QKF is required to compute the time and measurement update steps at each sampling instant and these two steps are derived in the sequel.

B. Time Update of the GS-QKF

Assume at time k the Gaussian sum approximation of the the posterior density $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ is known and given by

$$\begin{aligned} p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) = \sum_{i=1}^{n} \alpha_{(k-1|k-1)i} \mathcal{N} \\ \times \left(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{(k-1|k-1)i}, P_{(k-1|k-1)i}\right). \end{aligned} \tag{58}$$

For a process noise model of (55), the transition prior $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ can be obtained as

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \sum_{j=1}^{l_1} \beta_{kj} \mathcal{N}\left(\mathbf{x}_k; f(\mathbf{x}_{k-1}) + \bar{\mathbf{v}}_{kj}, Q_{kj}\right). \tag{59}$$

[4]The EM algorithm learns the parameters of a stochastic model from a set of incomplete data. For example, in fitting the data points to a mixture model, the EM algorithm seeks to find the mixture model parameters having a large likelihood for the given data set. On the other hand, k-means algorithm takes an incremental approach to find the homogenous groups of data points known as clusters. It dynamically adds cluster-centers through a deterministic global search procedure. The MATLAB source code for these mixture learning algorithms can be downloaded from http://lear.inrealpes.fr/verbeek/software.

Using (8), the predicted density of $\mathbf{x}_k$ can then be written as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \sum_{j=1}^{l_1}\sum_{i=1}^{n}\alpha_{(k-1|k-1)i}\beta_{kj} \times \int_{\mathbb{R}^{nx}} \mathcal{N}\left(\mathbf{x};\hat{\mathbf{x}}_{(k|k-1)i}, P_{(k|k-1)i}\right) \times \mathcal{N}\left(\mathbf{x}_k; f(\mathbf{x}) + \bar{\mathbf{v}}_{kj}, Q_{kj}\right)d\mathbf{x}. \quad (60)$$

The integral on the right side is approximated by a Gaussian sum in $\mathbf{x}_k$ using the time update of the QKF. The predicted density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is therefore approximated as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \approx \sum_{j=1}^{l_1}\sum_{i=1}^{n}\alpha_{(k|k-1)ij} \times \mathcal{N}\left(\mathbf{x}_k;\hat{\mathbf{x}}_{(k|k-1)ij}, P_{(k|k-1)ij}\right) = \sum_{r=1}^{nl_1}\alpha_{(k|k-1)r}\mathcal{N}\left(\mathbf{x}_k;\hat{\mathbf{x}}_{(k|k-1)r}, P_{(k|k-1)r}\right). \quad (61)$$

Here, we have

$$\begin{aligned}
\alpha_{(k|k-1)r} &= \alpha_{(k-1|k-1)i}\beta_{kj}\\
\hat{\mathbf{x}}_{(k|k-1)r} &= \sum_{l=1}^{m}\omega_l X^*_{li,k|k-1} + \bar{\mathbf{v}}_{kj}\\
P_{(k|k-1)r} &= Q_{kj} + \sum_{l=1}^{m}\omega_l X^*_{li,k|k-1}X^{*T}_{li,k|k-1} - (\hat{\mathbf{x}}_{(k|k-1)r} - \bar{\mathbf{v}}_{kj})(\hat{\mathbf{x}}_{(k|k-1)r} - \bar{\mathbf{v}}_{kj})^T\\
X^*_{li,k|k-1} &= f\left(\sqrt{P_{(k-1|k-1)i}}\xi_l + \hat{\mathbf{x}}_{(k-1|k-1)i}, \mathbf{u}_{k-1}, k-1\right)\\
&\quad l = 1,\ldots,m.
\end{aligned}$$

C. Measurement Update of the GS-QKF

Assume that the measurement noise sequence is non-Gaussian and modeled by a Gaussian mixture as given by (56). The measurement likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$ can be written as

$$p(\mathbf{z}_k|\mathbf{x}_k) = \sum_{j=1}^{l_2}\mu_{kj}\mathcal{N}\left(\mathbf{z}_k; h(\mathbf{x}_k) + \bar{\mathbf{w}}_{kj}, R_{kj}\right). \quad (62)$$

The posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ can be approximated by a Gaussian sum after receiving the measurement $\mathbf{z}_k$ using the measurement update step of the new QKF:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{nl_1}\sum_{j=1}^{l_2}\alpha_{(k|k)r}\mathcal{N}\left(\mathbf{x}_k;\hat{\mathbf{x}}_{(k|k)r}, P_{(k|k)r}\right) = \sum_{r=1}^{nl_1l_2}\alpha_{(k|k)r}\mathcal{N}\left(\mathbf{x}_k;\hat{\mathbf{x}}_{(k|k)r}, P_{(k|k)r}\right). \quad (63)$$

Here

$$\begin{aligned}
\hat{\mathbf{x}}_{(k|k)r} &= \hat{\mathbf{x}}_{(k|k-1)i} + W_{kr}\left(\mathbf{z}_k - \hat{\mathbf{z}}_{(k|k-1)r}\right)\\
P_{(k|k)r} &= P_{(k|k-1)i} - W_{kr}P_{zz,r}W^T_{kr}\\
\alpha_{(k|k)r} &= \frac{\alpha_{(k|k-1)i}\Omega_{kr}\mu_{kj}}{\sum_{j=1}^{l_2}\sum_{i=1}^{nl_1}\alpha_{(k|k-1)i}\Omega_{kr}\mu_{kj}}
\end{aligned}$$

with

$$\begin{aligned}
W_{kr} &= P_{xz,r}P^{-1}_{zz,r}\\
P_{xz,r} &= \sum_{l=1}^{m}\omega_l X_{l(k|k-1)i}Z^T_{l(k|k-1)i} - \hat{\mathbf{x}}_{(k|k-1)i}(\hat{\mathbf{z}}_{(k|k-1)r} - \bar{\mathbf{w}}_{kj})^T\\
P_{zz,r} &= R_{kj} + \sum_{l=1}^{m}\omega_l Z_{l(k|k-1)i}Z^T_{l(k|k-1)i} - (\hat{\mathbf{z}}_{(k|k-1)r} - \bar{\mathbf{w}}_{kj})(\hat{\mathbf{z}}_{(k|k-1)r} - \bar{\mathbf{w}}_{kj})^T\\
\hat{\mathbf{z}}_{(k|k-1)r} &= \sum_{l=1}^{m}\omega_l Z_{l(k|k-1)i} + \bar{\mathbf{w}}_{kj}\\
Z_{l(k|k-1)i} &= h\left(X_{l(k|k-1)i}, \mathbf{u}_k, k\right)\\
X_{l(k|k-1)i} &= \sqrt{P_{(k|k-1)i}}\xi_l + \hat{\mathbf{x}}_{(k|k-1)i} \quad l = 1,2,\ldots,m\\
\Omega_{kr} &= \mathcal{N}\left(\mathbf{z}_k;\hat{\mathbf{z}}_{(k|k-1)r}, P_{zz,r}\right).
\end{aligned}$$

The final state estimate $\hat{\mathbf{x}}_{k|k}$ in the minimum-mean-squared-error sense and the associated covariance $P_{k|k}$ are computed using the following theorem [37]:

Theorem: Let $p(\mathbf{x})$ be the density function of a normalized mixture of n Gaussians such that $p(\mathbf{x}) = \sum_{i=1}^{n}\omega_i\mathcal{N}(\mu_i, \Sigma_i)$. Let $p_A(\mathbf{x}) = \mathcal{N}(\mu, \Sigma)$ be a Gaussian density defined as

$$\mu = \sum_{i=1}^{n}\omega_i\mu_i \quad (64)$$

$$\Sigma = \sum_{i=1}^{n}\omega_i\left(\Sigma_i + [\mu_i - \mu][\mu_i - \mu]^T\right). \quad (65)$$

Then $p_A(\mathbf{x})$ has first two moments (mean and covariance) as $p(\mathbf{x})$. Furthermore, $p_A(\mathbf{x})$ minimizes the Kullback-Leibler divergence between $p(\mathbf{x})$ and any Gaussian density.

From (63), observe that the number of Gaussian terms has increased from n to nl_1l_2 at the end of the measurement update step of the GS-QKF due to Gaussian-sum approximation of the process and measurement noise. Consequently, the growth in the number of terms at each time step will reduce the practicability of this approximation. The next section is devoted to various Gaussian mixture reduction techniques that can be used to solve this issue.

VIII. GAUSSIAN MIXTURE REDUCTION

There are three classes of methods for dealing with large-scale Gaussian mixture reduction (GMR) techniques [23], [55]—i) decision directed methods where the number of Gaussian terms are pruned according to specific decision rules; ii) randomized pruning which includes methods such as probabilistic teacher; and iii) quasi-Bayesian approximation. Since it is required to reduce dynamically a large number of Gaussian sum components into a small number, this paper adopts quasi-Bayesian approximation. It includes three different techniques. They are: i) pruning; ii) joining; and iii) integral squared error-based GMR.

A. Pruning

Pruning means discarding Gaussian mixture components with negligible probability weights and keeping the remaining ones in a Gaussian sum density. This can be accomplished by adopting any one of the following methods:

- retaining Gaussian mixture components with the largest weights and discarding the remaining ones;
- discarding the set of Gaussian mixture components with the smallest weights such that the total weights discarded do not exceed a threshold ϵ;
- retaining all components with weights greater than or equal to a threshold ϵ and discarding Gaussian mixture components with weights less than ϵ, where ϵ is chosen depending on the problem at hand.

B. Joining

In this method, the distance measure used to gauge the similarity of two Gaussian mixture components i and j is based on the Mahalonobis distance d_{ij} [49]:

$$d_{ij}^2 = \frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)^T \Sigma^{-1} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j). \tag{66}$$

Here, Σ is the combined covariance for the entire Gaussian sum density, α_i and α_j are the means of the weights and $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_j$ are the i-th and j-th Gaussian mixture components. From (66), it can be inferred that this algorithm favors merging Gaussian mixture components carrying lower probability weights over those carrying higher probability weights. The assumption under merging is that the mean and covariance of the original Gaussian mixture density should be preserved. In light of this assumption, Gaussian mixture components with the lowest distance can be merged as shown below [49], [61].

$$\begin{aligned}
\text{weight}: \quad & \alpha_c = \alpha_i + \alpha_j \\
\text{mean}: \quad & \bar{\mathbf{x}}_c = \frac{1}{\alpha_i + \alpha_j} \{\alpha_i \bar{\mathbf{x}}_i + \alpha_j \bar{\mathbf{x}}_j\} \\
\text{Covariance}: \quad & \Sigma_c = \frac{1}{\alpha_i + \alpha_j} \left\{ \alpha_i \Sigma_i + \alpha_j \Sigma_j + \frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j} \right. \\
& \left. \times (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)^T \right\}.
\end{aligned}$$

C. Integral Squared Error-Based Gaussian Mixture Reduction

Pruning and joining techniques reduce the Gaussian mixture components by considering the individual pairs of Gaussian mixture. In contrast, the Gaussian mixture reduction (GMR) in the minimum-integral-squared error sense takes into account the full density and minimizes the integral squared error-based cost function [61].

$$\Psi_r = \arg\min \int_{\mathbb{R}^{n_x}} (p(\mathbf{x}; \Psi_h) - p(\mathbf{x}; \Psi_r))^2 d\mathbf{x}. \tag{67}$$

Here $p(\mathbf{x}; \Psi_h)$ denotes the original Gaussian mixture with parameters Ψ_h and $p(\mathbf{x}; \Psi_r)$ is the density with a reduced number of components and parameters given by Ψ_r. The GMR in the minimum-integral-squared error sense utilizes (67) to determine if merging or pruning action is to be performed on Gaussian mixture components so that the chosen action will tend to produce a small increase in the cost function. Since both the original density and the potential reduced order approximation of this density are Gaussian mixtures, we can express both the cost function and its derivative in a closed form. Consequently, a gradient descent-based optimization strategy can be utilized to improve the results. A detailed derivation of this method can be found in [61].

Remarks:

- The Gaussian sum filter algorithm is derived assuming "small" covariances [2]. However, the covariance of each Gaussian sum term grows, especially when the covariance of the process noise

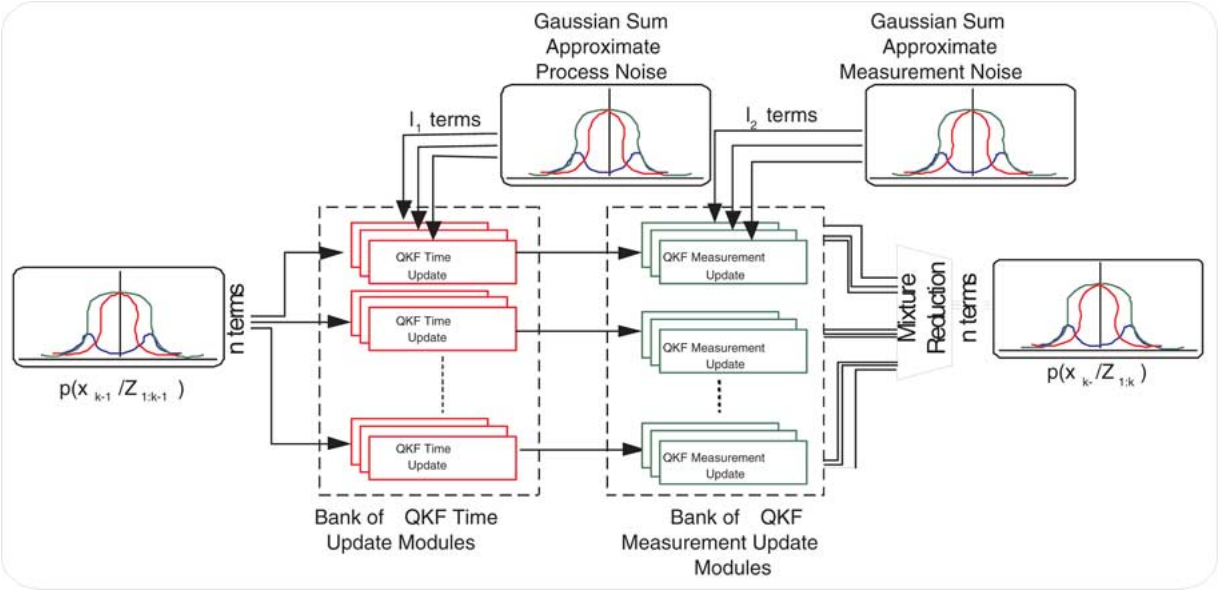

Fig. 2. ***Flow Diagram of the GS-QKF.***

is large. To combat this problem, it has been suggested that we approximate the Gaussian process noise as a finite sum of Gaussians [2].

- The number of terms of the Gaussian sum which are retained at each iteration is guided by the problem at hand. This criterion can be chosen as a compromise between computational complexity and performance gain.

IX. GS-QKF FILTER STRUCTURE

This section summarizes the structure of the proposed GS-QKF. The block diagram of the new GS-QKF is presented in Fig. 2. It consists of five stages.

Stage 1: Previous Estimate

Assume at time k that the Gaussian sum approximate posterior density $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) = \sum_{i=1}^{n} \alpha_{(k-1|k-1)i} \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{(k-1|k-1)i}, P_{(k-1|k-1)i})$ is known.

Stage 2: Time Update

The posterior density $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ from stage 1 and the Gaussian sum approximate process noise $\mathbf{v}_k \sim \sum_{i=1}^{l_1} \beta_{ki} \mathcal{N}(\mathbf{v}_k; \bar{\mathbf{v}}_{ki}, Q_{ki})$ are fed to the bank of $n \times l_1$ QKF time update modules. At the end of this stage, the predicted density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \approx \sum_{r=1}^{nl_1} \alpha_{(k|k-1)r} \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{(k|k-1)r}, P_{(k|k-1)r})$ is obtained as in (61).

Stage 3: Measurement Update

The predicted density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ from stage 2 and the Gaussian sum approximate measurement noise $\mathbf{w}_k \sim \sum_{i=1}^{l_2} \mu_{ki} \mathcal{N}(\mathbf{w}_k; \bar{\mathbf{w}}_{ki}, R_{ki})$ are fed to the bank of $n \times l_1 \times l_2$ QKF measurement update modules. At the end of this stage, the posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{r=1}^{nl_1l_2} \alpha_{(k|k)r} \times \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{(k|k)r}, P_{(k|k)r})$ is obtained as in (63).

Stage 4: Output of the State Estimate

The state estimate $\hat{\mathbf{x}}_{k|k}$ in the minimum-mean-squared-error sense and the associated covariance $P_{k|k}$ are estimated from the Gaussian sum approximate posterior density as in (64) and (65) respectively.

Stage 5: Mixture Reduction

Lastly, to control the number of Gaussian terms in the posterior density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, one of the proposed GMR techniques is applied.

The five stage procedure is repeated at each time step.

X. ACCURACY AND COMPUTATIONAL COMPLEXITY OF QUADRATURE POINT APPROACH

We can gain valuable insight into the QKF by analyzing the theoretical accuracy of the quadrature point scheme. For this purpose, consider a Gaussian-distributed scalar valued random variable x with mean $\bar{x}$ and variance Σ_x. We wish to calculate the mean $\bar{y}$ and the variance Σ_y of a random variable y which is related to x through the nonlinear function $y = g(x)$. As discussed in Section IV, the integral $\bar{y} = \int_{\mathbb{R}} g(x) \mathcal{N}(x; \bar{x}, \Sigma_x) dx$ is exact if $g(x)$ is a polynomial of order less than or equal to $(2m-1)$, where m is number of quadrature points. Similarly, $\Sigma_y = [\int_{\mathbb{R}} g^2(x) \mathcal{N}(x; \bar{x}, \Sigma_x) dx - \bar{y}^2]$ is exact if the order of the polynomial $g(x)$ is less than m.

For a polynomial of order greater than $(2m-1)$, the error, e associated with the m-point Gauss–Hermite quadrature approximation is given by [43], [46]:

$$e = \frac{m! g^{(2m)}(\varepsilon)}{(2m)!}. \tag{68}$$

Here $g^{(2m)}(\varepsilon)$ is the $2m$-th derivative of g evaluated at some appropriate (but unknown), where ε is in the same region as x, and the weight function is assumed to be the standard Gaussian density function. This error analysis can be extended readily to the multidimensional case as well. On the other hand, the linearization approach built into the EKF truncates the Taylor series expanded with respect to $\bar{\mathbf{x}}$ after the first term, and so it is only correct up to the first order of the Taylor series. A detailed error analysis of the EKF can be found in the appendix of [31].

From the above discussion, we see that the higher the number m of quadrature points (per axis) is, the lesser the estimation error becomes. However, the estimation accuracy comes with an increased computational complexity. In the QKF case, the evaluation of initial quadrature points and weights, $\{\omega_l, \xi_l\}_{l=1}^{m}$ requires the most computations; they are determined from a quadrature point set which approximates the univariate standard Gaussian density. These initial points and the associated weights can be determined off-line in advance. In terms of on-line computation, the QKF requires $O(m^{n_x})$ functional evaluations. The number of functional evaluations scales geometrically with the size of the state vector dimension. Fortunately, due to the increasing availability of cheap and powerful commercially-off-the-shelf (COTS) computational resources (e.g., general purpose CPUs, DSPs, etc.), on-line functional evaluations become viable even for a "large" dimensional state space model. For $m = 3$ and $n_x = 5$, for example, the QKF is required to perform some hundreds of functional evaluations which can be done on-line with currently available computational resources.

A. Numerical Example

Consider a random variable $\mathbf{z}_p = [r \;\; \theta]^T$ in polar coordinates, distributed as a Gaussian density described by $p(\mathbf{z}_p) = \mathcal{N}(\bar{\mathbf{z}}_p, \Sigma_p)$. The first component of this random variable r corresponds to the "range" and the second θ to the "angle". Let the mean range $\bar{r}$ be 80 units and the mean angle $\bar{\theta}$ be 0.61 radians, with $\Sigma_p = \text{diag}[60 \;\; 0.6]$. Then the random variable $\mathbf{z}_p$ in polar coordinates is converted to $\mathbf{z}_c = [x \; y]^T$ in Cartesian coordinates using the nonlinear transformation

$$\mathbf{z}_c = g(\mathbf{z}_p) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Our objective is to compare the statistics of the nonlinearly transformed random variable $\mathbf{z}_c$ obtained from different schemes. Fig. 3 shows the results of this transformation. The cloud of random samples represents the true density of the random variable $\mathbf{z}_c$. Observe that the transformed variable $\mathbf{z}_c$ is clearly non-Gaussian. This cloud is obtained by the Monte Carlo approach: 5000 samples were drawn from the Gaussian density $\mathcal{N}(\bar{\mathbf{z}}_p, \Sigma_p)$. Each sample was then propagated through the nonlinear transformation $g(.)$. The true mean and the covariance of $\mathbf{y}$ were computed from this sample cloud.

Essentially, three different linearization techniques were applied to capture the mean and the associated error covariance of the transformed variable $\mathbf{z}_c$. Firstly, the statistical linearization approach built into the QKF was applied via the linear regression through the quadrature (regression) points. The quadrature points were chosen such that they would capture the mean and covariance of the random variable $\mathbf{z}_p$. These points were then propagated through the nonlinear function $g(.)$. The mean and covariance of the random variable $\mathbf{y}$ were computed as described in Section IV.

Secondly, the set of sigma points built into the unscented Kalman filter (UKF) was propagated through the nonlinear function and the mean and covariance were computed as in [29].

Finally, the analytical linearization approach built into the EKF, was applied. It linearizes the nonlinear function $g(.)$ through a first-order Taylor series expansion around the mean $\bar{\mathbf{z}}_p$. The density of $\mathbf{z}_c$ is then approximated by a Gaussian density with mean $g(\bar{\mathbf{z}}_p)$ and the associated covariance $G\Sigma_p G^T$, where

$$G = \left[\nabla_{z_p} g(\mathbf{z}_p)\right]_{\mathbf{z}_p = \bar{\mathbf{z}}_p} = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta. \end{pmatrix}_{\theta=\bar{\theta}, r=\bar{r}.}$$

Fig. 3 illustrates the true mean and covariance (a 2σ-ellipse), of $\mathbf{z}_c$ together with the statistical linearization obtained via 3-point and 5-point quadrature scheme (per axis), and the analytical linearization approach. In this 2-D example, the true 2σ-ellipse should contain approximately 86% of random samples [48].

As can be seen in Fig. 3, the mean values calculated via the statistical linearization using the 3-point and 5-point quadrature scheme lie on top of the true mean whereas the the mean obtained by the analytical linearization (EKF) is biased with respect to the true mean. This can be attributed to the fact that the statistical linearization employs a nonzero bias correction term, "$\mathbf{b}$", whereas in the analytical linearization, the corresponding bias term $G(\mathbf{z}_p - \bar{\mathbf{z}}_p)$ is zero. Moreover, the 5-point quadrature scheme yields a more accurate well-conditioned (since it is less elongated), covariance estimate than the 3-point scheme, as expected. However, increasing the number of

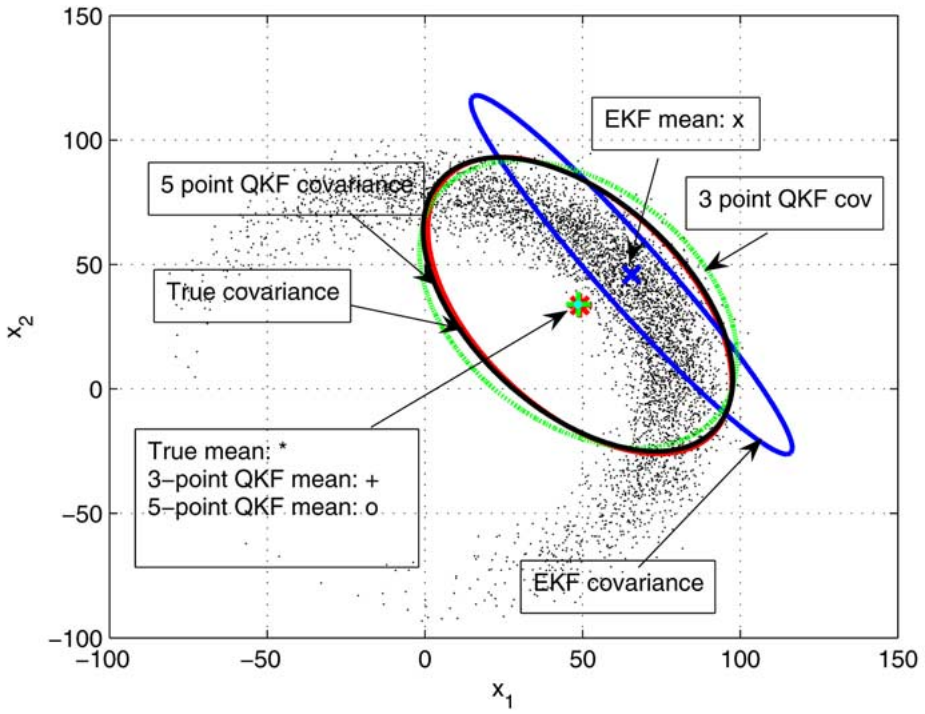

Fig. 3. ***The probability density function of a nonlinearly transformed Gaussian random variable.***

quadrature points beyond a certain limit at the expense of increased computational cost will not improve the filter performance significantly. The reason is that the QKF is a linear estimator which requires only the first two moments. In contrast, an increased number of quadrature points may improve the performance of the estimators which propagate higher order moments and this is yet to be investigated. In the considered example, $3^2(=9)$ points are required in the 3-point quadrature scheme, whereas the number of quadrature points is $5^2(=25)$ in the 5-point scheme. As shown in Fig. 3, the QKFs' contours lie on top of the true contour almost everywhere implying a more consistent estimate whereas, the linearization approach yields an inconsistent estimate. The contour plot of the sigma point scheme was seen to fall closely on top of the 3-point QKF and it is not shown in Fig. 3 for the clarity of the picture. Hence, it can be concluded that the 5-point quadrature scheme is more accurate than the sigma point scheme.

XI. FURTHER APPLICATIONS OF THE NEW QKF

In the previous section, we illustrated how the Gauss–Hermite rule can be used to estimate the statistics of a nonlinearly transformed Gaussian variable. Moreover, the consistency of the new QKF estimate was justified mathematically using the SLR theory in Section V. In addition, the new insight obtained from the development of the new QKF can be used to extend its applicability further in the following ways:

- The new QKF gives a more accurate least-squares solution: Conventionally, the nonlinear, least-squares problem is solved by the on-line Gauss–Newton method, which iteratively approximates the inverse of the empirical Fisher information matrix. The nonlinear least-squares problem can also be formulated as a parameter-estimation, and subsequently, the iterative EKF can be applied [6], [56]. Since it is possible to obtain a more accurate, statistically linearized measurement model, A_h, which can also be viewed as the statistically averaged Jacobian matrix of $h(\mathbf{x}_k)$, the least-squares problem can be solved efficiently using the measurement update of the new QKF.
- The new QKF gives a more accurate prediction in a continuous-time process model: for a continuous-time, nonlinear process model, an approximate closed form solution for the prediction covariance can be obtained if the nonlinear process model can be expressed as a linear model as closely as possible during the sampling time interval δ, [4]. In the case of a continuous-time, linear process model with process matrix $A(t)$ at time t, the one step

predicted error covariance matrix is given by the solution of the following Riccati equation [4]:

$$\frac{dP(t)}{dt} = A(t)P(t) + P(t)A^T(t) + Q(t). \quad (69)$$

The matrix differential equation in (69) can be solved by assuming that $A(.)$ remains constant for the time interval $[k, k+\delta]$ so that

$$\begin{aligned} P_{k|k+\delta} &= \exp(A\delta)P_{k|k}\exp(A^T\delta) + \delta Q(k) \\ &\approx \psi(k, k+\delta)P_{k|k}\psi^T(k, k+\delta) + \delta Q(k) \end{aligned} \quad (70)$$

where

$$\psi(k, k+\delta) = I_{n_x} + \delta A(k) + \frac{1}{2}\delta^2 A(k)^2. \quad (71)$$

In the case of the new QKF framework, $A(k)$ can be replaced by the statistically linearized process matrix $A_{f,k}$ in order to obtain more accurate predicted error covariance matrix.

- The linear regression allows the estimation of a nonlinear system with discontinuous functions: since the new QKF requires only the functional evaluations and not the derivatives of the $f(.)$ and $h(.)$, it can be applied to nonsmooth, nonanalytical systems (e.g., a saturation function).
- The new QKF algorithm allows the estimation in correlated or nonadditive noise: The new QKF allows us to extend its applicability to correlated, or nonadditive Gaussian process and measurement noise by augmenting the state vector and the associated covariance. Such structures commonly arise in algorithms such as the Schmidt-Kalman filter and UKF.

Since the new QKF shares a number of similarities such as point-wise evaluation of nonlinearities and weighted sum statistics, it can be considered as a new member of the family of sigma point filters. In the Gaussian quadrature scheme, we choose the quadrature points in one dimension and use them to create a grid of points in $\mathbb{R}^{n_x}$. On the other hand, the sigma points are chosen by taking into account the whole spread of the distribution. The plain quadrature points in the 3-point per axis scheme lie at the corners and the mid-points of a hyper-cube with fixed dimension, whereas the plain sigma points lie at the intersection of the axes and the surface of a hyper-sphere with fixed radius except a point at the center in both schemes. Hence, the number of quadrature points scales geometrically with the state vector dimension, whereas the number of sigma points increases linearly. It is noteworthy here that the 3-point quadrature point set is exactly identical to the sigma point set in a 1-D state space. In the multidimensional case, the calculation of the mean and covariance of a nonlinearly transformed variable, using the 3-point quadrature scheme, will be exact if the highest order of the nonlinear function is less than or equal to five, whereas it is correctly to the third order in the case of sigma points. Another major difference between these two schemes is that the weights in the quadrature point scheme are positive and the unit sum with the obvious advantage of SLR interpretation, whereas the weights can be negative in the sigma point filters. However, if we view the nonlinear filtering problem as a numerical integration problem, the quadrature and sigma point sets are two instances of the numerical integration technique.

XII. EXPERIMENTAL RESULTS

We present three examples to compare the performance of the QKF and the GS-QKF with currently used methods, namely the standard sampling-importance resampling particle filter (SIR-PF), the EKF, and the GS-EKF. The first example is a univariate nonstationary growth model with additive Gaussian noise. The second example considers the same model as in the first example, but with heavy-tailed additive gamma distributed process noise. The third is a practical, target tracking problem in glint noise over a 4-D state space. Glint is a random wandering of the apparent measured position of a target due to reflections from different elements of the target. Consequently, it affects the radar measurement by producing a heavy-tailed non-Gaussian distribution. The glint noise is approximated by various non-Gaussian distributions (e.g., a Gaussian mixture model [26]).

A. One-Dimensional Nonlinear Example With Additive Gaussian Noise

Consider the scalar nonlinear system described by

$$x_k = \phi_1 x_{k-1} + \phi_2 x_{k-1}^2 + 8\cos(\omega k) + v_k$$

where v_k is zero-mean Gaussian white noise with variance 10; $\omega = 1.2$, $\phi_1 = 0.2$, and $\phi_2 = 0.01$ are scalar parameters. The state x_k is to be estimated from the measurement data z_k.

$$z_k = x_k^2 + w_k$$

where w_k is zero-mean Gaussian white noise with variance 0.01. This example is nonlinear, both in the process and the measurement equation and a similar example is also considered in [21] for the performance evaluation of

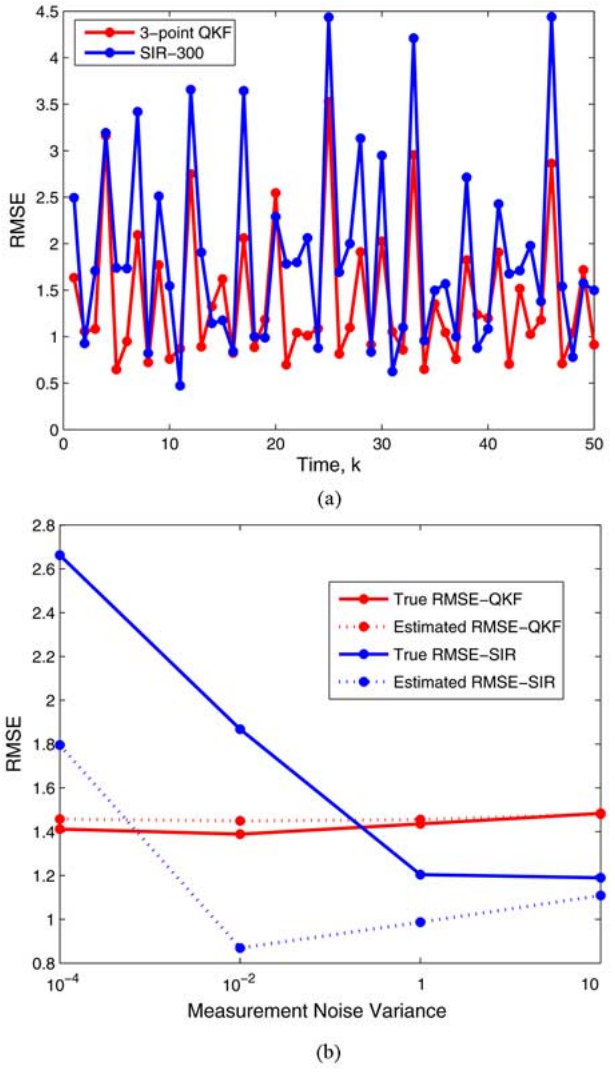

Fig. 4. ***One-dimensional nonlinear example with additive Gaussian noise. (a) Root mean square error (RMSE) of the 3-point QKF and the SIR-PF with 300 particles. (b) True RMSE and the estimated RMSE of the QKF and the SIR particle filter with 300 particles for different measurement noise levels.***

various filters. The experiment was repeated 100 times. In each run, the actual initial state x_0 is assumed to be a uniformly distributed random variable in the interval $[0, 1]$. The prior state $x_{0|0}$ at time 0 is assumed to be Gaussian such that $x_{0|0} \sim \mathcal{N}(0.5, 2)$. The number of quadrature points is chosen to be 3. The SIR-PF uses 300 particles. Fig. 4(a) shows the root-mean-squared-error (RMSE) of the QKF and the SIR-PF. The performance of the EKF is not shown in Fig. 4(a), as it often yields a divergent estimate.

The reason why the SIR-PF performs badly in this problem is due to the highly peaked measurement-likelihood function arising from relatively small noise variance. The SIR-PF uses the transition prior as the proposal distribution and, consequently, samples of the transition prior remote from the measurement likelihood are effectively wasted. However, the SIR-PF can improve its performance by moving the particles towards the highly likelihood region with the help of a linearized Kalman filter [56]. We also vary the measurement noise variance from 10^{-4} to 10 and the true RMSE and the filter estimated RMSE or the square-root of the error variance are obtained as shown in Fig. 4(b). Observe from Fig. 4(b) that the true RMSE of the SIR-PF is too high when the measuring device is more accurate or the measurement noise level is too small. This is counterintuitive. Moreover, the SIR-PF underestimates its own error at small noise levels, and consequently, the discrepancy is more significant. However, the SIR-PF works well when the sensor possesses the "right" amount of noise. Fig. 4(b) shows that the QKF estimated RMSE matches well with its overall true RMSE implying that the QKF yields a consistent estimate irrespective of the measurement noise level.

B. One-Dimensional Nonlinear Non-Gaussian Example

We consider the same problem as described in the first example, but this time the process noise v_k is assumed to be distributed as a heavy-tailed gamma function and given by $v_k \sim \Gamma(3, 2)$, where Γ denotes the gamma distribution. The observation noise w_k is assumed to be zero-mean Gaussian with variance 0.01. The experiment was repeated 100 times with a random initialization in each run as described in the first example. The SIR-PF uses 500 particles. We compare the performance of the GS-QKF against GSEKF and SIR-PF. Two different GS-QKFs are considered: first, the GS-QKF (5-1-1) uses a 5-component Gaussian mixture model (GMM) for the state posterior, and a 1-component GMM for both the process and measurement noise densities. Secondly, the GS-QKF (5-3-1) uses a 5-component GMM for state posterior, a 3-component GMM to approximate the "heavy-tailed" gamma distributed process noise (see Fig. 5) and a 1-component GMM for the measurement noise density. For this example, all three GMR techniques described in Section VIII perform almost equally well, and we do not provide a performance comparison among the GMR techniques. Fig. 6 shows the RMSE of the different filters when pruning is employed as a GMR technique in Gaussian sum filters.

A factor responsible for the bad performance of the SIR-PF in this problem can be attributed to the same reason for the bad performance of the SIR-PF as mentioned in the first example. In addition to this, the support of the proposal distribution, or the gamma distribution in this case, may not be broad enough to cover the actual posterior and this may be another factor for the performance degradation of the SIR-PF. The GS-QKF (5-3-1) clearly outperforms the GS-QKF (5-1-1). The superior performance of the GS-QKF

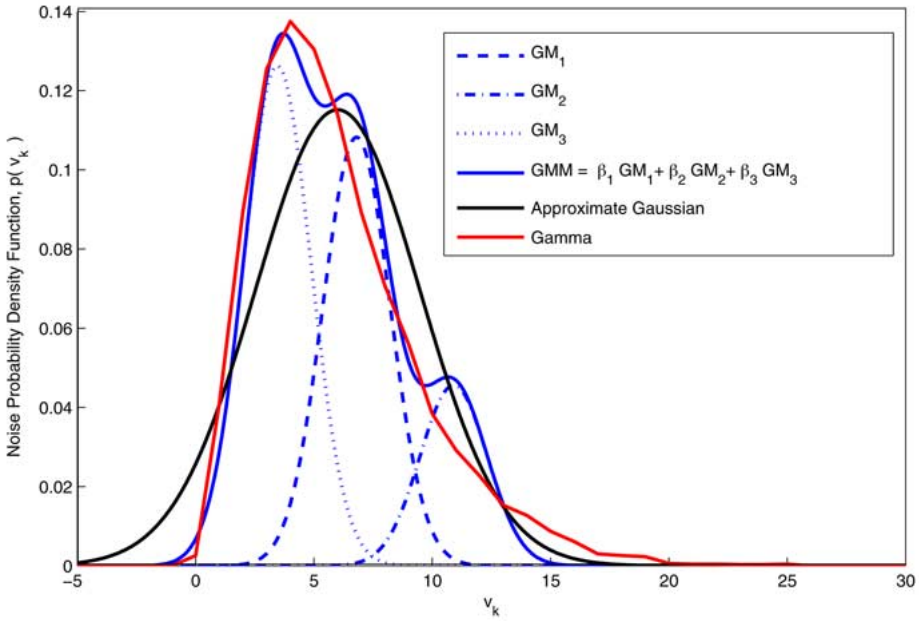

Fig. 5. ***Gaussian mixture model (GMM) approximation of heavy-tailed, asymmetric gamma distribution. The red curve is the gamma distribution and the black curve is the single Gaussian approximation of the gamma distribution. The dotted-line, blue curves show the shape and position of the three Gaussian mixture (GM) components of a GMM approximation shown by solid, blue line. The GMM was fitted to the gamma distribution using the expectation-maximization (EM) algorithm.***

(5-3-1) can be attributed to the fact that it models the non-Gaussian nature of the process noise better than that of the GS-QKF (5-1-1). We also observe from Fig. 6 that the GS-EKF performs well without diverging. The reason is that the Gaussian sum approximate process noise has mixture terms with sufficiently low noise variances, thereby validating the EKF's linearization assumption in a small region of the state space.

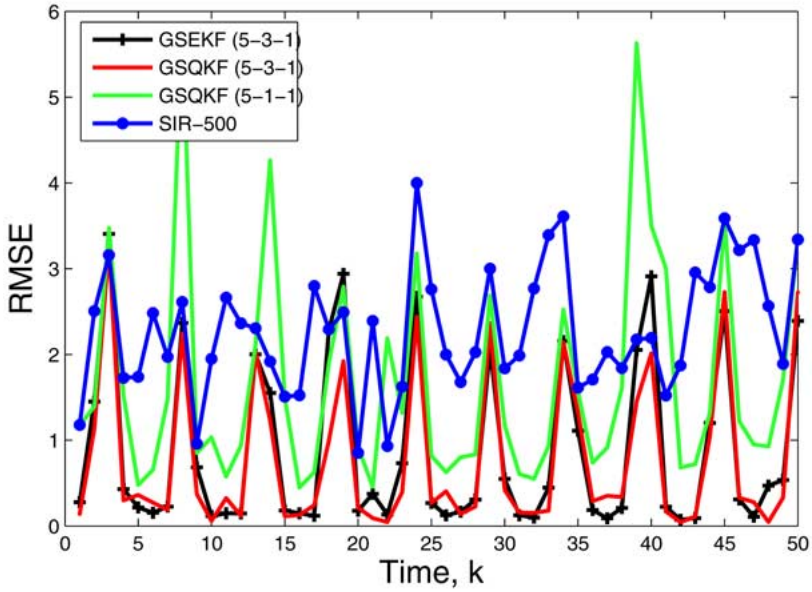

Fig. 6. ***RMSE of the GS-EKF (5-3-1), GS-QKF (5-3-1), GS-QKF (5-1-1) and SIR particle filter with 500 particles.***

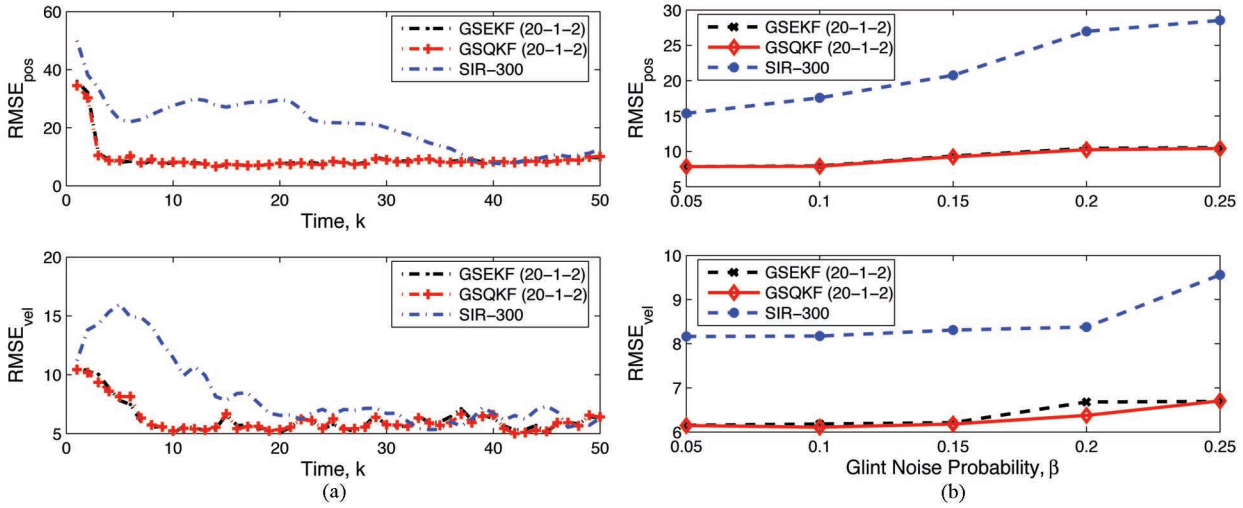

Fig. 7. *Target tracking with glint noise. (a) RMSE of position and velocity of GS-EKF (20-1-2), GS-QKF (20-1-2), and SIR particle filter with 300 particles for the glint noise probability $\epsilon_w = 0.15$. (b) RMSE of position and velocity of the GS-EKF (20-1-2), GS-QKF (20-1-2), and SIR particle filter with 300 particles for various glint noise probabilities.*

C. Target Tracking With Glint Noise

In this last example, the new GS-QKF is applied to the problem of radar target tracking with glint noise and its performance is compared with the GS-EKF and the SIR-PF with 300 particles. The following state-space model describes the tracking scenario [5]:

$$\mathbf{x}_k = \begin{pmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{x}_{k-1} + \begin{pmatrix} \Delta^2/2 & 0 \\ \Delta & 0 \\ 0 & \Delta^2/2 \\ 0 & \Delta \end{pmatrix} \mathbf{v}_k \quad (72)$$

$$\mathbf{z}_k = \begin{pmatrix} \sqrt{x_k^2 + y_k^2} \\ \tan^{-1}\left(\frac{y_k}{x_k}\right) \end{pmatrix} + \mathbf{w}_k. \quad (73)$$

Here $\mathbf{x}_k = [x_k \ \dot{x}_k \ y_k \ \dot{y}_k]^T$ with x and y denote the cartesian coordinates of the target with velocities $\dot{x}$ and $\dot{y}$ in x and y directions; the sampling interval Δ is set to be 0.5 s; the process noise sequence $\mathbf{v}_k$ is zero mean Gaussian with covariance $Q_k = \text{diag}([50\ \text{m}^2\text{s}^{-4} \ \ 5\ \text{m}^2\text{s}^{-4}])$; the glint measurement noise, $\mathbf{w}_k$ is modeled by a Gaussian mixture as $p(\mathbf{v}_k) = (1-\beta)\mathcal{N}(\mathbf{0}, R_1) + \beta\mathcal{N}(\mathbf{0}, R_2)$, where $R_1 = \text{diag}([50\ \text{m}^2 \ \ 0.4\ \text{mrad}^2])$ and $R_2 = \text{diag}([5000\ \text{m}^2 \ \ 16\ \text{mrad}^2])$ and β refers to the glint noise probability; the actual initial state is defined to be $\mathbf{x}_0 = [10\,000\ \text{m} \ \ 300\ \text{ms}^{-1} \ \ 1000\ \text{m} \ \ -40\ \text{ms}^{-1}]^T$; the initial estimate $\hat{\mathbf{x}}_{0|0}$ and the associated covariance $P_{0|0}$ are assumed to be $[10\,175\ \text{m} \ \ 295\ \text{ms}^{-1} \ \ 980\ \text{m} - 35 \ \ \text{ms}^{-1}]^T$ and $\text{diag}([100^2\ \text{m}^2 \ \ 10^2\ \text{m}^2\text{s}^{-2} \ \ 100^2\ \text{m}^2 \ \ 10^2\ \text{m}^2\text{s}^{-2}])$ respectively. The number of Gaussian terms used in the Gaussian sum filters is 20 at the end of each iteration. The *a priori* density $p(\mathbf{x}_0) = \mathcal{N}(\hat{\mathbf{x}}_{0|0}, P_{0|0})$ is also approximated by a Gaussian sum of 20 components. We assume that the radar is placed at the origin.

For performance evaluation, the RMSE in the position and velocity are obtained in 100 trials. Fig. 7(a) shows the RMSE for a glint noise probability $\beta = 0.15$. From Fig. 7(a), it is observed that the GS-EKF performs similarly to the GS-QKF. The reason is that the nonlinearity is limited to the measurement model only. However, the SIR-PF performs badly. This is because the initial covariance is so "large" that it takes long time to converge. It should be noted that in the experiment described herein, the computational complexity of the particle filter exceeds that of the GS-EKF and GS-QKF. Moreover, the use of an increased number of particles and/or the advanced particle filtering techniques could further improve the performance of the SIR-PF but that would be at the expense of increased complexity and computational burden.

The tracking performance of position and velocity as a function of glint noise probability is presented in Fig. 7(b). A close inspection of Fig. 7(b) reveals that the GS-QKF, in general, outperforms the GS-EKF for any value of glint probability.

XIII. SUMMARY AND DISCUSSION

The celebrated Kalman filter is the optimal solution for linear Gaussian filtering problems. For many nonlinear problems, the extended Kalman filter (EKF) has

successfully been applied. Nevertheless, the EKF does breakdown when the reference point, say the prior mean, is far from the posterior mean. The motivation for this paper is to obtain an improved solution to nonlinear problems in the Kalman filtering framework. We present a systematic formulation of the so-called quadrature Kalman filter (QKF) from the SLR perspective. The regression points are obtained from the Gauss–Hermite quadrature scheme which approximates a Gaussian density discretely. We also explored how this new insight can be used to enhance the performance of the QKF for a given application. However, the QKF suffers from the curse of dimensionality; this problem becomes more severe for a high dimensional problem, especially when the state vector size goes beyond six.

Since the QKF assumes a Gaussian posterior assumption, it may fail in certain nonlinear non-Gaussian problems with heavily-tailed or multimodal posterior. For nonlinear, non-Gaussian problems, a well-known suboptimal algorithm is the sequential Monte Carlo (SMC) filter. Although, the SMC filter is rooted in the Bayesian framework, it is an approximate simulation-based algorithm. In this paper, we extend the application of the QKF to nonlinear non-Gaussian problems by incorporating the conventional Gaussian sum filters. The resulting filtering algorithm, called the Gaussian sum QKF, approximates both the predicted and posterior densities as Gaussian sums. The limitation with this approach is that the number of Gaussian terms grows exponentially over time. In order to alleviate the growth of Gaussian sum components, three different Gaussian mixture reduction techniques are presented. Due to the use of the QKF as a subcomponent, the GS-QKF also suffers from similar shortcomings as the QKF. For example, the statistical linear approximation technique in a highly nonlinear filtering problem may be insufficient, thereby leading to divergent state estimate of the QKF; for the same filtering problem but with additive non-Gaussian noise case, the GS-QKF is also likely to yield a divergent state estimate. As an alternative, one could possibly use the SMC method that has the advantage of having no explicit functional assumption in this case.

Our future research will focus on the following issues:

1) Exploring extensions of and refinements to the new nonlinear filtering algorithm, including the use of square-root filtering for improved computational accuracy.
2) As already mentioned, tackling nonlinear filtering problems of high dimensional complexity using the new algorithms, the curse of dimensionality could be a serious practical issue; we will explore the application of supervised learning and cubature rules as tools for the approximate computation of the known nonlinear functions that feature prominently in the algorithm.
3) To test the performance of the new algorithms, we will undertake the experimental study of a challenging real-life problem and compare the performance against the Monte-Carlo methods currently in use. ■

Acknowledgment

The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for its financial support. They would also like to thank Dr. S. Godsill, Dr. V. Krishnamurthy, and Dr. H. Kushner for their valuable input.

REFERENCES

[1] L. Aggoun and R. J. Elliott, *Measure Theory and Filtering*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[2] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Control*, vol. AC-17, no. 4, pp. 439–448, Aug. 1972.

[3] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[4] M. Athans, R. P. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noise measurements," *IEEE Trans. Autom. Control*, vol. 13, pp. 504–514, 1968.

[5] Y. B. Shalom, X.-R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. New York: Wiley, 2001.

[6] B. M. Bell and F. W. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Trans. Autom. Control*, vol. 38, no. 2, pp. 294–297, Feb. 1993.

[7] D. P. Bertsekas, "Incremental least-square methods and the extended Kalman filters," *SIAM J. Optim.*, vol. 6, no. 3, pp. 807–822, Aug. 1996.

[8] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*. New York: Academic, 1977.

[9] I. Bilik and J. Tabrikian, "Target tracking in Glint noise environment using nonlinear non-Gaussian Kalman filter," in *Proc. 2006 IEEE Int. Radar Conf.*, pp. 282–286.

[10] R. S. Bucy and K. D. Senne, "Digital synthesis of nonlinear filters," *Automatica*, vol. 24, no. 6, pp. 789–801, 1974.

[11] R. S. Bucy and H. Youssef, "Nonlinear filter representation via spline functions," in *Proc. 5th Symp. Nonlinear Estimation*, 1974, pp. 51–60.

[12] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*. New York: Springer, 2005.

[13] O. Cappe, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proc. IEEE*, vol. 95, no. 4, Apr. 2007.

[14] S. Challa, Y. Bar-Shalom, and V. Krishnamurthy, "Nonlinear filtering via generalized Edgeworth series and Gauss Hermite Quadrature," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1816–1820, Jun. 2000.

[15] R. J. P. de Figueiredo and J. G. Jan, "Spline filters," in *Proc. 2nd Symp. Nonlinear Estimation*, 1971, pp. 127–138.

[16] A. Doucet, J. de Freitas, and N. Gordon, *Sequential Monte Carlo in Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[17] R. J. Elliott, *Stochastic Calculus and Applications*. New York: Springer-Verlag, 1982.

[18] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.

[19] G. H. Golub and J. H. Welsch, "Calculation of Gauss Quadrature rules," *Math. Comput.*, vol. 23, no. 106, pp. 221–230, 1969.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[21] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc. F*, vol. 140, pp. 107–113, Apr. 1993.

[22] I. Gyongy, "Approximations of stochastic partial differential equations," in *Stochastic Partial Differential Equations*. New York: Dekker, 2002, vol. 277, *Lecture Notes in Pure and Applied Mathematics*, pp. 287–307.

[23] P. J. Harrison and C. F. Stevens, "Bayesian forecasting," *J. Royal Stat. Soc.*, ser. B, vol. 38, no. 3, pp. 205–247, 1976.

[24] S. Haykin, K. Huber, and Z. Chen, "Bayesian sequential state estimation for MIMO wireless communications," *Proc. IEEE*, vol. 92, no. 3, pp. 439–455, Mar. 2004.

[25] S. Haykin, P. Yee, and E. Derbez, "Optimum nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2774–2786, Nov. 1997.

[26] G. A. Hewer, R. D. Martin, and J. Zeh, "Robust preprocessing for Kalman filtering of glint noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 120–128, Jan. 1987.

[27] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[28] A. Jazwinski, *Stochastic Processing and Filtering Theory*. New York: Academic, 1970.

[29] S. J. Julier, "The scaled unscented transform," in *Proc. Amer. Control Conf.*, 2000, pp. 4555–4559.

[30] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[31] S. J. Julier, J. K. Ulhmann, and H. F. 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. 472–482, Mar. 2000.

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

[33] S. C. Kramer and H. W. Sorenson, "Recursive Bayesian estimation using piecewise constant approximation," *Automatica*, vol. 24, no. 6, pp. 789–801, 1988.

[34] V. Krishnan, *Probability and Random Process*. New York: Wiley, 2006.

[35] H. J. Kushner and A. S. Budhiraja, "A nonlinear filtering algorithm based on an approximation of the conditional distribution," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 580–585, Mar. 2000.

[36] H. J. Kushner, "Dynamic equations for optimal nonlinear filtering," *J. Differ. Eq.*, vol. 3, pp. 179–190, 1971.

[37] S. Lauriitzen, *Graphical Models*. New York: Oxford Univ. Press, 1996.

[38] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Comment on 'A new method for the nonlinear transformation of means and covariances in filters and estimators'," *IEEE Trans. Autom. Control*, vol. 47, no. 8, pp. 1406–1408, Aug. 2002.

[39] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Kalman filters for non-linear systems: A comparison of performance," *Int. J. Control*, vol. 77, no. 7, pp. 639–653, May 2004.

[40] D. G. Lainiotis and J. G. Desppande, "Parameter estimation using splines," *Inf. Sci.*, vol. 7, pp. 291–315, 1974.

[41] D. Middleton, "Man-made noise in urban environments and transportaion systems," *IEEE Trans. Commun.*, vol. COM-21, no. 11, pp. 1232–1241, Nov. 1973.

[42] L. Mo, X. Song, Y. Zhou, Z. Sun, and Y. Bar-Shalom, "Unbiased converted measurements in tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 1023–1027, Jul. 1998.

[43] J. C. Naylor and A. F. M. Smith, "Applications of a method for the efficient computation of posterior distributions," *Appl. Stat.*, vol. 31, no. 3, pp. 214–225, 1982.

[44] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation of nonlinear systems," *Automatica*, vol. 36, pp. 1627–1638, 2000.

[45] B. Oksendal, *Stochastic Differential Equations*, 5th ed. Berlin, Germany: Springer, 1998.

[46] R. J. Phaneuf, "Approximate nonlinear estimation," Ph.D dissertation, Mass. Inst. Technol., Cambridge, 1968.

[47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[48] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond Kalman Filters: Particle Filters for Applications*. Norwood, MA: Artech House, 2004.

[49] D. Salmond, "Mixture reduction algorithms for target tracking in clutter," *SPIE Signal and Data Processing of Small Targets*, vol. 1305, pp. 434–445, Apr. 1990.

[50] T. S. Schei, "A finite difference method for linearizing in nonlinear estimation algorithms," *Automatica*, vol. 33, no. 11, pp. 2051–2058, 1997.

[51] A. F. M. Smith, A. M. Skene, J. E. H. Shaw, and J. C. Naylor, "Progress with numerical and graphical methods for practical Bayesian statistics," *Statistician*, vol. 36, pp. 75–82, 1982.

[52] H. W. Sorenson, "Recursive estimation for nonlinear dyanmic systems," in *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall, Ed. New York: Marcel Dekker Inc., 1988, ch. 6.

[53] H. W. Sorenson and A. R. Stubberud, "Nonlinear filtering by approximation of *a posteriori* density," *Int. J. Control*, vol. 8, no. 1, pp. 33–51, 1968.

[54] K. Srinivasan, "State estimation by orthogonal expansion of probability distribution," *IEEE Trans. Automat. Control*, vol. AC-15, no. 1, pp. 3–10, Feb. 1970.

[55] D. M. Titterington, D. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. Chichester, U.K.: Wiley, 1985.

[56] R. van der Merwe. (2004). "Sigma-point filters for probabilistic inference in dynamic state space models," Ph.D dissertation, OGI School Sci. Eng., Beaverton, OR. [Online]. Available: http://www.cse.ogi.edu/rudmerwe/pubs/index.htm

[57] J. J. Verbeek, J. R. J. Nunnink, and N. Vlassis, "Accelerated EM-based clustering of large data sets," *Data Mining Knowl. Discovery*, vol. 13, no. 3, pp. 291–307, 2006.

[58] A. H. Wang and R. L. Klein, "Implementation of nonlinear estimation using monospline," in *Proc. 13th IEEE Conf. Decision and Control*, 1976, pp. 1305–1307.

[59] A. H. Wang and R. L. Klein, "Optimal quadrature formula nonlinear estimations," *Inf. Sci.*, vol. 16, pp. 169–184, 1978.

[60] M. West and J. Harrison, *Bayesian Forecasting and Dynamical Models*, 2nd ed. New York: Springer-Verlag, 1989.

[61] J. L. Williams. (2003). "Gaussian mixture reduction for tracking multiple maneuvering targets in clutter," M.S.E.E. thesis, Air Force Inst. Technol., Wright-Patterson Air Force Base, OH. [Online]. Available: http://ssg.mit.edu/jlwil/

[62] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer-Verlag, 1985.

[63] M. Zakai, "On the optimal filtering of diffusion processes," *Wahrschein likehkeitstheorrie verw gebiete*, vol. 11, pp. 230–243, 1969.

ABOUT THE AUTHORS

Ienkaran Arasaratnam was born in Jaffna, Sri Lanka, in 1978. He received the B.Sc.Eng. degree with first class honors in electronic and telecomm engineering from the University of Moratuwa, Sri Lanka, in 2003 and the M.A.Sc degree with outstanding thesis research award in electrical engineering from the McMaster University, Hamilton, ON, Canada, in 2006. He is currently working toward the Ph.D. degree at the same university.

His research interests include the development of efficient nonlinear inference algorithms and their applications to auditory signal processing and target tracking.

Simon Haykin (Fellow, IEEE) received the B.Sc. (First-class Honours), Ph.D., and D.Sc. degrees in electrical engineering from the University of Birmingham, U.K.

Currently, he holds the title Distinguished University Professor in the ECE Department at McMaster University, Hamilton, ON, Canada. He is the author of numerous books, including the most widely used books: *Communication Systems* (4th ed., Wiley), *Adaptive Filter Theory* (4th ed., Prentice-Hall), *Neural Networks: A Comprehensive Foundation* (2nd ed., Prentice-Hall) and the newly published book *Adaptive Radar Signal Processing* (Wiley), as well as numerous refereed journal papers.

Prof. Haykin is a Fellow of the Royal Society of Canada, recipient of the Honourary Degree of Doctor of Technical Sciences from ETH, Zurich, Switzerland, and the Henry Booker Gold Medal from URSI, as well as other prizes and awards.

Robert J. Elliott received the B.S. and M.S. degrees from Oxford University, U.K., and the Ph.D. and D.Sc. degrees from Cambridge University.

He has held positions at Newcastle, Yale, Oxford, Warwick, Hull, Alberta, and visiting positions in Toronto, Northwestern, Kentucky, Brown, Paris, Denmark, Hong Kong, and Australia. Currently he is the RBC Financial Group Professor of Finance at the University of Calgary, Calgary, AB, Canada, where he is also an Adjunct Professor in both the Department of Mathematics and the Department of Electrical Engineering. He is also an Adjunct Professor in Electrical Engineering at McMaster University, Hamilton, ON, Canada, and in Mathematics at the University of Adelaide and the Australian National University. He also consults for banks and energy companies. He has authored nine books and over 350 papers. His book with P. E. Kopp, *Mathematics of Financial Markets*, was published by Springer in 1999. The Hungarian edition was published in 2000 and an expanded second Edition appeared in 2004. Springer-Verlag published his book *Binomial Methods in Finance*, written with J. van der Hoek, in 2006. He has also worked in signal processing, and his book with L. Aggoun and J. Moore, *Hidden Markov Models: Estimation and Control*, was published in 1995 by Springer Verlag and reprinted in 1997. His book with L. Aggoun, *Measure and Filtering* was published by Cambridge University Press in June 2004.