A Numerical-Integration Perspective on Gaussian Filters

Yuanxin Wu, Dewen Hu, *Senior Member, IEEE*, Meiping Wu, and Xiaoping Hu

***Abstract*—This paper proposes a numerical-integration perspective on the Gaussian filters. A Gaussian filter is approximation of the Bayesian inference with the Gaussian posterior probability density assumption being valid. There exists a variation of Gaussian filters in the literature that derived themselves from very different backgrounds. From the numerical-integration viewpoint, various versions of Gaussian filters are only distinctive from each other in their specific treatments of approximating the multiple statistical integrations. A common base is provided for the first time to analyze and compare Gaussian filters with respect to accuracy, efficiency and stability factor. This study is expected to facilitate the selection of appropriate Gaussian filters in practice and to help design more efficient filters by employing better numerical integration methods.**

***Index Terms*—Cramer–Rao bound, Gaussian filter, monomial, nonlinear filtering, numerical-integration, product rule, stability factor.**

I. Introduction

THE role of filtering in the practical systems—from spacecraft attitude estimation [1] to ballistic target tracking [2], [3] and precision agriculture [4], [5]—cannot be overemphasized at all. Almost all real systems involve nonlinearity of one kind or another to which the Bayesian inference provides an optimal solution framework for dynamic state estimation problems [6]. Because the Bayesian solution requires the propagation of the full probability density, in general the optimal nonlinear filtering is analytically intractable. Approximations are thus necessary, e.g., Gaussian approximation to the posterior probability density [7]–[13] or approximating the density with piecewise constant functions [14]. The class of filters derived under Gaussian assumption is commonly called the Gaussian filters. So far, there has been a variation of Gaussian filters that derived themselves from very different backgrounds. The most celebrated one is the extended Kalman filter (EKF) [15], which came into being as a result of adapting any encountered nonlinear system to the linear Kalman filter [9] by approximating the nonlinearity through successive linearization at the current estimate. Since its first successful application in the Apollo guidance system in 1968 [15], [16], the EKF has been applied to deal with nonlinear filtering problems in many practical systems. When being employed to address significant nonlinearities, however, the EKF may become very difficult to tune and even prone to divergence. Long-term experiences have shown that it is only reliable for systems that are almost linear in the update interval. Most of the EKF's difficulties owe much to the local linearization at a single point in the probabilistic state space. There are other filters employing the high-order Taylor expansion, such as the second-order filter [17], which is not frequently used in engineering because it has to evaluate the cumbersome Hessian matrix.

Recently, an unscented transformation (UT)-based Kalman filter [18] arose and was widely approved as a promising substitute for the EKF [19]–[21]. The UT was developed to address the deficiencies of local linearization using an intuitive method of transforming statistical information.[1] In light of the intuition that approximating a probability distribution is easier than approximating an arbitrary nonlinear transformation, the UT makes probabilistic inference by 1) parameterizing the mean and covariance of a probability distribution via a set of deterministically selected samples, 2) propagating them through the true nonlinear transformation, and 3) calculating the parameters of the propagated Gaussian approximation from the transformed samples. Eliminating cumbersome derivation and evaluation of Jacobian/Hessian matrices, the UT-based unscented Kalman filter (UKF) is much easier to implement and outperforms the EKF. Lefebvre *et al.* [22] interpreted the UT as a statistical linear regression (a discrete implementation of the statistical linearization [17], which, in contrast to the local linearization, employs the system information at multiple points in the state space). This insight justifies from another aspect the derivative-free UKF's benefits over the local linearization-based EKF.

Ito *et al.* [7] proposed two Gaussian filters from the standpoint of numerical integration. The so-called Gauss–Hermite filter[2] (GHF) was obtained by employing the product Gauss-Hermite quadrature rule while the central difference filter (CDF) was constructed using the polynomial interpolation methods. These two filters performed better than or as good as the UKF in numerical tests there. Meanwhile, Norgaard *et al.* [10] devised a divided difference filter (DDF) through approximating the derivatives, e.g., Jacobian/Hessian matrices, by the central divided differences formed using the Sterling's polynomial interpolation

Manuscript received October 27, 2004; revised August 16, 2005. This work was supported in part by the National Natural Science Foundation of China (60374006, 60234030, and 30370416), the Distinguished Young Scholars Fund of China (60225015), and the Ministry of Education of China (TRAPOYT Project). The associate editor coordinating the review of this paper and approving it for publication was Prof. Tulay Adali.

The authors are with the Laboratory of Inertial Technology, Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan, China, 410073 (e-mail: yuanx_wu@hotmail.com; dwhu@nudt.edu.cn; meipingwu@263.net; xiaopinghu@hotmail.com).

Digital Object Identifier 10.1109/TSP.2006.875389

[1]The motivation of the UT was recently well documented in a survey paper [21].

[2]It should be reminded that the GHF was once presented several decades ago [23], [24].

formula. Recently, the Gaussian particle filter (GPF) was suggested making use of the idea of Bayesian sampling [8]. With the Gaussian assumption being valid, the GPF is asymptotically optimal in the number of random samples, which means that equipped with the computational ability to handle a large number of samples, the GPF is supposed to outperform any analytical Gaussian filter. Kotecha and Djuric [8] have done a nice comparison among the GPF, the EKF, and the UKF. As pointed out in [25], the GPF actually extends the analytical Gaussian filters using Monte Carlo integration and the Bayesian update rule. In view of its asymptotical optimality, it may be promising to use the GPF as a guide telling how best a Gaussian filter could perform, just like the role of the Cramer–Rao bound [26] for any nonlinear filter. The disadvantage of any random sampling-based filter, such as the GPF, is that we have to assume great computational burden, usually hundreds of times larger than an analytical filter does. Therefore, the discussions below exclude the GPF and are dedicated to analytical Gaussian filters. It is worthy to remember a practical rule: only when all analytical tools do not suffice the requirements is the simulation-based method necessary.

A question now arises: with so many different Gaussian filters, how do we decide which one is suitable for a filtering problem in hand? In the literature, there have been efforts to compare part of the above-mentioned Gaussian filters on a case-by-case base [2], [27]. However, it is not advisable to implement and compare all filters before making the final decision. Different from previous literature [7], [10], [15], [17], [18], this paper is intended not to come up with novel filters but to build a common platform to analyze and compare all off-the-shelf Gaussian filters with no or restricted need of numerical simulations.[3] Hopefully, we wish to obtain from this study some constructive insights that help design novel Gaussian filters in future research.

The contents are organized as follows. Section II gives a concise formulation of the general Gaussian filter. Section III derives all of the existing approximate Gaussian filters through applying different methods of numerical integration to calculate the multidimensional statistical integrals necessitated in the general Gaussian filter. Section IV makes a comparative study of the approximate Gaussian filters with respect to accuracy, efficiency and stability factor. In numerical examples, we also carry out the Cramer–Rao lower bound for the reference of performance comparison. Discussions and conclusions are in Section V.

II. General Gaussian Filter Formulation

Now we formulate the general Gaussian filter in a concise manner. For a rigorous development, readers are referred to [7] and [29] for details. Consider a discrete-time nonlinear system written in the form of dynamic state-space model as

$$\begin{aligned} x_n &= f(x_{n-1}) + w_{n-1} \quad \text{(process equation)} \\ y_n &= h(x_n) + v_n \quad \text{(observation equation)} \end{aligned} \tag{1}$$

[3]Quite recently, Lefebvre *et al.* [28] made a less comprehensive review on Kalman filters than this paper from the standpoint of statistical linearization.

where $f(\cdot)$ and $h(\cdot)$ are some known functions. The random process noise $w_{n-1} \sim \mathcal{N}(0, Q)$ is uncorrelated with the system states at time instant k ($\leq n-1$); the random measurement noise $v_n \sim \mathcal{N}(0, R)$ and is uncorrelated with the system state and the process noise at all time instants.

The nonlinear filtering problem is to estimate the true but unknown system state x_n given the noisy observations at time instant k ($\leq n$), denoted by $x_{n|n}$ in the sequel. If the probability density is well approximated by the Gaussian distribution, a Gaussian filter in the Kalman-like structure (using linear update rule) can be used to address the estimation task [7], [20], [21]. The Gaussian distribution being uniquely characterized by its first two moments (mean and covariance), the general Gaussian filter is formulated as [7], [29, pp. 249–260]

$$\begin{aligned} x_{n|n} &= x_{n|n-1} + K_n(y_n - \hat{y}_n) \\ P_{n|n} &= P_{n|n-1} - K_n P_{y_n y_n} K_n^T \\ K_n &= P_{x_n y_n} P_{y_n y_n}^{-1} \end{aligned} \tag{2}$$

where

$$\begin{aligned} x_{n|n-1} &= E\left[f(x_{n-1})\right] \\ P_{n|n-1} &= E\left[(x_n - x_{n|n-1})(x_n - x_{n|n-1})^T\right] + Q \\ \hat{y}_n &= E\left[h(x_n)\right] \\ P_{y_n y_n} &= E\left[(y_n - \hat{y}_n)(y_n - \hat{y}_n)^T\right] + R \\ P_{x_n y_n} &= E\left[(x_n - x_{n|n-1})(y_n - \hat{y}_n)^T\right]. \end{aligned} \tag{3}$$

It is clear that the general Gaussian filter above necessitates five expectations as in (3). For example, the optimal prediction $x_{n|n-1}$, conditioned on the observations at time instant k ($\leq n-1$), corresponds to the expectation of the nonlinear process function of the random variables x_{n-1}, taken over the posterior probability density of the state at time instant $n-1$, i.e.,

$$\begin{aligned} x_{n|n-1} = E\left[f(x_{n-1})\right] = \int_{\mathbb{R}^d} f(t) \frac{1}{\left((2\pi)^d \det P_{n-1|n-1}\right)^{1/2}} \\ \times \exp\left(-\frac{1}{2}(t - x_{n-1|n-1})^T P_{n-1|n-1}^{-1}(t - x_{n-1|n-1})\right) dt \end{aligned} \tag{4}$$

where d is the dimension of the system state. The second expectation is taken over the posterior probability density at time instant $n-1$ as well while the last three expectations are taken over the prior probability density at time instant n. Hence, these expectations involve five multiple integrals, two for mean and three for covariance, of the same dimension as the state.

III. Approximate Gaussian Filters

Now for brevity we consider a multiple integral of the form

$$\int_{\mathbb{R}^d} F(x) \frac{1}{\left((2\pi)^d \det P\right)^{1/2}} \exp\left(-\frac{1}{2}(x - \overline{x})^T P^{-1}(x - \overline{x})\right) dx \tag{5}$$

where x is a random variable with mean $\overline{x}$ and covariance P. Numerically approximating (5), an integral over an infinite integration region with a Gaussian weight, is also an important issue in finances and statistics, etc. [30], [31].

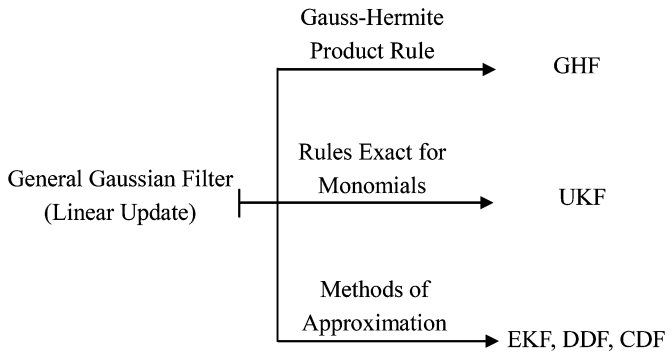


Fig. 1. Relations among the Gaussian filters.

Let $P = S^T S$. Since covariance P is positive, it is always possible to find such a matrix square root S, e.g., using the efficient Cholesky decomposition [32]. Substituting $x = \overline{x} + S^T t$ into the right-hand side of (5) yields

$$\begin{aligned}\int_{\mathbb{R}^d} F(x) &\frac{1}{((2\pi)^d \det P)^q 1/2} \exp\left(-\frac{1}{2}(x-\overline{x})^T P^{-1}(qx-\overline{x})\right) dx \\ &\overset{\substack{x=\overline{x}+S^T t \\ P=S^T S}}{=} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} F(\overline{x}+S^T t) \\ &\quad \times \exp\left(-\frac{1}{2} t^T S (S^T S)^{-1} S^T t\right) dt \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} F(\overline{x}+S^T t) \exp\left(-\frac{1}{2} t^T t\right) dt \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{F}(t) \exp\left(-\frac{1}{2} t^T t\right) dt \\ &\triangleq I[\tilde{F}] \end{aligned} \tag{6}$$

with $\tilde{F}(t) = F(\overline{x} + S^T t)$. The above is a normal Gaussian weighted integral over the unbounded $\mathbb{R}^d$ space. For a general nonlinear function $F(x)$, the above integral cannot be solved analytically. Various approximations to (6) result in a variation of approximate Gaussian filters in the present literature, as noted in [7]. The connection between the general Gaussian filter and the existing approximate Gaussian filters is sketched in Fig. 1. Comprehensive discussions will be made case by case in the immediate sequel.

Generally speaking, most multiple integration rules are designed to integrate a certain class of polynomials exactly. Before getting involved in further details, we first introduce the definition of the *precision* of an integration rule [33].

Definition: A rule is said to have precision p if it integrates monomials up to degree p exactly, that is, monomials $\prod_{i=1}^d x_i^{k_i}$ with $k_i \geq 0$ and $\sum_{i=1}^d k_i \leq p$, but not exactly for some monomial of degree $\sum_{i=1}^d k_i = p+1$.

A. Product Gauss–Hermite Rule

The most natural approach to address a multiple integral is to treat it as a nested sequence of the univariate integral and then to use the unidimensional quadrature rule with respect to each argument in turn. The resulting multiple integration formula is a *product rule* [33], [34]. Regarding the multiple integral in (6), the Gauss–Hermite quadrature rule can be used [32]. It says that, given an integer m, we can find a set of points $u_{1,2,\ldots,m}$ and weights $w_{1,2,\ldots,m}$ such that the approximation

$$\frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} g(x) \exp\left(-\frac{1}{2}x^2\right) dx = \sum_{i=0}^{m-1} w_i g(u_i) \tag{7}$$

is exact for polynomials of degree $2m-1$, i.e., the linear combination of monomials up to degree $2m-1$. The Gauss–Hermite rule is best in the sense that it integrates exactly monomials of as high degree as possible and thus excellent for large classes of functions arising in practice. The program for generating weights and points is readily off the shelf [32]. Given $m = 3$, for example

$$u_0 = 0,\ u_1 = -\sqrt{3},\ u_2 = \sqrt{3} \text{ and } w_0 = \frac{2}{3},\ w_1 = w_2 = \frac{1}{6}. \tag{8}$$

By the product rule, (6) is approximated as

$$\begin{aligned} I_{\text{GHF}}[\tilde{F}] &= \sum_{i_0,i_1,\ldots,i_{d-1}=0}^{m-1} w_{i_0} w_{i_1} \cdots w_{i_{d-1}} \tilde{F}(u) \\ &= \sum_{i_0,i_1,\ldots,i_{d-1}=0}^{m-1} w_{i_0} w_{i_1} \cdots w_{i_{d-1}} F(\overline{x}+S^T u) \end{aligned} \tag{9}$$

where $u = [u_{i_0}, u_{i_1}, \ldots, u_{i_{d-1}}]^T$. The product rule based on the m-point Gauss–Hermite quadrature integrates exactly monomials $\Pi_{i=1}^d x_i^{k_i}$ with $k_i \leq 2m-1$, but it does not integrate exactly x_i^{2m} and is therefore of precision $2m-1$.

Applying the above product Gauss–Hermite rule to calculate the five integrals in (3) gives birth to the GHF [7], [23], [24]. The filter badly suffers from *the curse of dimensionality*, that is to say, the number of function evaluations m^d grows exponentially with the state dimension d. Depending on the complexity of the integrand (process and observation functions), computation expenses due to a large number of function evaluations usually limit the use of the GHF to practical systems of dimension no more than five or six.

B. Rules Exact for Monomials

By allowing nonproduct rules, it is possible to find efficient rules in the sense of having precision $2m-1$ yet requiring function evaluations fewer than m^d [33]–[37]. The points of such a rule are chosen directly in $\mathbb{R}^d$ instead of using a grid of points.[4] Recall that the integral in (6) is not ordinary but has a special property in that both the integral region and the weight function are *fully symmetric*. More formally [35], a set Δ is said to be fully symmetric if $x \in \Delta$ implies $y \in \Delta$, where y is obtained from x by permutations and changes of sign of the coordinates of x. A function $g(\cdot)$ defined on a fully symmetric set is fully symmetric if $g(x) = g(y)$. It follows that any finite fully symmetric set can be decomposed into a finite number of disjoint subsets having the property that any member of a particular subset can be used to generate the whole subset. In the

[4]Novak *et al.* [38] presented an interpolatory method for high-dimensional integration using the Smolyak formulas, which are linear combinations of product formulas with the special properties, i.e., only products with a relatively small number of points are used and the linear combination is chosen such that the interpolation property for is preserved for $d = 1$ is preserved for $d > 1$.

sequel, the set generated by a point $[u_1, u_2, \ldots, u_k, 0, \ldots, 0]^T$ in $\mathbb{R}^d$ is denoted by $\langle u_1, u_2, \ldots, u_k \rangle$, where the zero coordinates have been suppressed for brevity. For example, given a point $x = [1, 0]^T$, then the set generated by x is $\langle 1 \rangle = \{[1, 0]^T, [-1, 0]^T, [0, -1]^T, [0, 1]^T\}$. The particular point is referred to as a *generator*.

Since the region $\mathbb{R}^d$ and the Gaussian weight function are both fully symmetric, the fully symmetric integration rule developed in [35] can be used to handle the integral in (6).

$$I[\tilde{F}] = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{F}(x) \exp\left(-\frac{1}{2}x^T x\right) dx = \sum w_i \tilde{F}\langle \cdot \rangle_i \tag{10}$$

where $\tilde{F}\langle \cdot \rangle_i$ is a short notation of $\sum_{x \in \langle \cdot \rangle_i} \tilde{F}(x)$ with the sum extending over all the points in the i^{th} fully symmetric set $\langle \cdot \rangle_i$. From the definition of fully symmetry, it can be readily proved that the fully symmetric integration rule is exact for any monomial odd in one or more coordinates.

For example, the rule of precision 3 (with $2d + 1$ function evaluations) takes the form

$$\begin{aligned} I[\tilde{F}] &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{F}(x) \exp\left(-\frac{1}{2}x^T x\right) dx \\ &= w_0 \tilde{F}\langle 0 \rangle + w_1 \tilde{F}\langle u_1 \rangle \end{aligned} \tag{11}$$

The set $\langle u_1 \rangle$ has $2d$ points while the set $\langle 0 \rangle$ contains only the origin. In such a case, we need only consider the monomials up to degree 3 without any odd power, i.e., 1 and x_1^2. Substituting the two monomials into (11) respectively yields

$$\begin{aligned} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} 1 \cdot \exp\left(-\frac{1}{2}x^T x\right) dx &= 1 = w_0 + w_1 \cdot 2d \\ \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} x_1^2 \exp\left(-\frac{1}{2}x^T x\right) dx &= 1 = w_1 \cdot 2u_1^2. \end{aligned} \tag{12}$$

The above equations[5] give $w_1 = 1/2u_1^2$ and $w_0 = 1 - (d/u_1^2)$, in which $u_1 (\neq 0)$ is a free parameter. With this rule, (6) is approximated as

$$\begin{aligned} I_{\mathrm{UKF3}}[\tilde{F}] &= w_0 \tilde{F}\langle 0 \rangle + w_1 \tilde{F}\langle u_1 \rangle \\ &= w_0 F(\bar{x}) + w_1 \sum_{i=1}^{d} F(\bar{x} + S^T u_1 e_i) \\ &\quad + w_1 \sum_{i=1}^{d} F(\bar{x} - S^T u_1 e_i) \end{aligned} \tag{13}$$

where e_i is the d-dimensional unit vector, with the ith coordinate being 1 and other coordinates 0. Applying the above rule to (3) yields right the UKF [18], [21]. This relationship was first revealed by Lerner [40]. In the UKF context, the free parameter u_1 was suggested to choose $\sqrt{3}$ so as to capture the fourth-order principal moment x_i^4.

It is possible to construct rules of higher precision along the same lines by solving a set of nonlinear equations in real domain. McNamee and Stenger [35] developed rules of arbitrary precision $2k + 1$ in $\mathbb{R}^d$ with the number of function evaluations being $\mathrm{O}((2d)^k/k!)$, $d \to \infty$. Capstick *et al.* [39] argued that the McNamee–Stenger scheme breaks down for some considerably high precision, e.g., 15, due to singular nonlinear equations, but it does not make a difference in this paper since we have no interests in such a high precision here. The McNamee–Stenger rule of precision 5 takes the form [40]

$$\begin{aligned} I_{\mathrm{UKF5}}[\tilde{F}] &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{F}(x) \exp\left(-\frac{1}{2}x^T x\right) dx \\ &= w_0 \tilde{F}\langle 0 \rangle + w_1 \tilde{F}\langle u_1 \rangle + w_2 \tilde{F}\langle u_1, u_1 \rangle. \end{aligned} \tag{14}$$

Now only the monomials 1, x_1^2, $x_1^2 x_2^2$, and x_1^4 need to be considered. By solving the resulting four nonlinear equations, we obtain

$$u_1 = \sqrt{3} \text{ and } w_0 = 1 + \frac{d^2 - 7d}{18},\ w_1 = \frac{4 - d}{18},\ w_2 = \frac{1}{36}. \tag{15}$$

Applying the rule of precision 5 to (3) gives birth to a higher order UKF, which is slightly different from the Julier's high-order UKF (see [21, App. IV]) in that the latter corresponds to a rule of precision 5 using the generators $\langle 0 \rangle$, $\langle u_1 \rangle$, and $\langle u_2, u_2 \rangle$. Julier's high-order UKF needs to solve four equations in five unknowns. The problem was addressed there by incorporating the sixth-order moments information. However, with a simple algebra manipulation, we can see that this rule exists only if $d \leq 4$ (otherwise, the nonlinear equations have no real roots; see footnote 5 for explanation). The UKF using a set of simplex sigma points [41], [42], [21, App. III] can also be readily derived from the integration rule of precision 2 [34, pp. 79–88].

In [37], Genz *et al.* constructed a fully symmetric interpolatory integration rule for multidimensional integrals in the form of (6). Unlike the McNamee–Stenger scheme, Genz's rule takes an explicit form and needs not to solve a set of nonlinear functions, i.e.,

$$I[\tilde{F}] = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{F}(x) \exp\left(-\frac{1}{2}x^T x\right) dx = \sum_{\mathbf{p} \in P^{m,d}} w_{\mathbf{p}} \tilde{F}\langle \lambda_{\mathbf{p}} \rangle. \tag{16}$$

Here, $\lambda_{\mathbf{p}} = (\lambda_{p_1}, \lambda_{p_2}, \ldots, \lambda_{p_d})$ in which λ_i are nonnegative real numbers with $\lambda_0 = 0$. $P^{(m,d)}$ is a set of all distinct d-partitions of the integers $0, 1, \ldots, m$ defined by

$$\begin{aligned} P^{(m,d)} = \{(p_1, p_2, \ldots, p_d) : m \geq p_1 \\ \geq p_2 \geq \ldots p_d \geq 0, |\mathbf{p}| \leq m\} \end{aligned} \tag{17}$$

with $|\mathbf{p}| = \sum_{i=1}^{d} p_i$. If the weights $w_{\mathbf{p}}$ are given by

$$w_{\mathbf{p}} = 2^{-K} \sum_{|\mathbf{k}| \leq m - |\mathbf{p}|} \prod_{i=1}^{d} \frac{a_{k_i + p_i}}{\prod_{j=0, \neq p_i}^{k_i + p_i} \left(\lambda_{p_i}^2 - \lambda_j^2\right)} \tag{18}$$

where K is the number of nonzero components in $\mathbf{p}$, and

$$a_i = \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}x^2\right) \prod_{j=0}^{i-1} \left(x^2 - \lambda_j^2\right) dx \tag{19}$$

[5]The desired equations are linear in weights, but nonlinear in points (see, for example, (12)). The solution can thus be extremely difficult to obtain and the points and/or weights may be complex, or there may even be no solution [37], [39]!

for $i > 0$, with $a_0 = 1$, then the integration rule has precision $2m+1$. To be more efficient, λ_i are determined by successive extensions of the one-dimensional 3-point Gauss–Hermite rule so that certain sets of weights vanish. For precision 3 ($m = 1$) and 5 ($m = 2$), we emphasize that Genz's rules are exactly the same as the counterparts of the McNamee–Stenger scheme [43]. The development is involved and omitted here. For precision no less than 7, Genz's rules diverge from and have a smaller stability factor than the corresponding McNamee–Stenger rules [39]. For the definition of the stability factor, readers are referred to Section IV.

C. Methods of Approximation

The idea is to approximate the integrand by a sequence of functions so that each of them can be integrated in closed form. The integral of the approximation then constitutes an approximation to the integral [33]. A common choice is to employ the Taylor polynomials. For example, the first-order Taylor polynomial approximation is

$$\tilde{F}(t) \approx \tilde{F}(0) + \tilde{F}'(0)t. \tag{20}$$

Substituting (20) into (6), we obtain

$$\begin{aligned} I_{\mathrm{EKF}}[\tilde{F}] &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \left(\tilde{F}(0) + \tilde{F}'(0)x\right) \exp\left(-\frac{1}{2}x^T x\right) dx \\ &= \tilde{F}(0). \end{aligned} \tag{21}$$

Obviously, this rule is of precision 1. It should be reminded that (21) can only be applied to the two mean integrals in (3). For covariance approximation, we need to consider an integral of the other form

$$\begin{aligned} &\int_{\mathbb{R}^d} \left(F_1(x) - I[\tilde{F}_1]\right)\left(F_2(x) - I[\tilde{F}_2]\right)^T \frac{1}{((2\pi)^d \det P)^{1/2}} \\ &\quad \times \exp\left(-\frac{1}{2}(x-\overline{x})^T P^{-1}(x-\overline{x})\right) dx \\ &\quad = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} F_1(\overline{x} + S^T t) F_2^T(\overline{x} + S^T t) \\ &\qquad \times \exp\left(-\frac{1}{2}t^T t\right) dt - I[\tilde{F}_1] I[\tilde{F}_2]^T \\ &\quad = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{F}_1(t)\tilde{F}_2^T(t) \exp\left(-\frac{1}{2}t^T t\right) dx - I[\tilde{F}_1] I[\tilde{F}_2]^T \\ &\quad = I\left[\tilde{F}_1 \tilde{F}_2^T\right] - I[\tilde{F}_1] I[\tilde{F}_2]^T \\ &\quad \triangleq J[\tilde{F}_1, \tilde{F}_2]. \end{aligned} \tag{22}$$

Different from I in (6), here J has three different Gaussian weighted integrands, namely, $\tilde{F}_1$, $\tilde{F}_2$, and their product $\tilde{F}_1 \cdot \tilde{F}_2^T$. The above definition of precision is not directly applicable to such a case, so we need a redefinition so as to encompass (22). *Specifically, we say a rule is of precision p if it is exact for each element monomial of the integrand up to degree p, but not exact for some element monomial of degree $p+1$.* Regarding (22), the integrand is the product $\tilde{F}_1(x)\tilde{F}_2^T(x)$, of which the factors $\tilde{F}_1(x)$, $\tilde{F}_2(x)$ are restricted to be monomials. For instance, if $\tilde{F}_1(x)\tilde{F}_2^T(x) = x_1^2$, then $\tilde{F}_1(x)$, $\tilde{F}_2(x)$ can be 1 and x_1^2 or x_1 and x_1, respectively.

Substituting (20) into (22)

$$\begin{aligned} &J_{\mathrm{EKF}}[\tilde{F}_1, \tilde{F}_2] \\ &\quad = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \left(\tilde{F}_1(0) + \tilde{F}_1'(0)x\right)\left(\tilde{F}_2(0) + \tilde{F}_2'(0)x\right)^T \\ &\qquad \times \exp\left(-\frac{1}{2}x^T x\right) dx - \tilde{F}_1(0)\tilde{F}_2^T(0) \\ &\quad = \tilde{F}_1'(0)\tilde{F}_2'^T(0). \end{aligned} \tag{23}$$

The above rule has precision 3. For example, let $\tilde{F}_1(x) = [1, x_1, x_1^2]^T$ and $\tilde{F}_2(x) = [1, x_1^2, x_2]^T$, then

$$\tilde{F}_1(x)\tilde{F}_2^T(x) = \begin{bmatrix} 1 & x_1^2 & x_2 \\ x_1 & x_1^3 & x_1 x_2 \\ x_1^2 & x_1^4 & x_1^2 x_2 \end{bmatrix}$$

and

$$\begin{aligned} J[\tilde{F}_1, \tilde{F}_2] &= I\left[\tilde{F}_1 \tilde{F}_2^T\right] - I[\tilde{F}_1] I[\tilde{F}_2]^T \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 3 & 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} [1\ 1\ 0] \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \\ J_{\mathrm{EKF}}[\tilde{F}_1, \tilde{F}_2] &= \tilde{F}_1'(0)\tilde{F}_2'^T(0) \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 2x_1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2x_1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\Bigg|_{x=0} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{24}$$

which indicates that J_{EKF} is exact for monomials of degree up to 3 but not for a monomial of degree 4, x_1^4 with $\tilde{F}_1(x) = \tilde{F}_2(x) = x_1^2$. It can be verified that J_{EKF} is exact for all monomials odd in one or more coordinates. Applying (21) and (23) to respectively approximate mean and covariance in (3) recovers the EKF. The high-order EKF can be similarly derived based on the corresponding high-order Taylor polynomial [17]. It is apparent from (21) and (23) that the EKF and its variations only employ the information at a single point in the probabilistic state space and have to evaluate the Jacobian/Hessian matrices, which is nontrivial and even impossible for cumbersome systems. An alternative is to replace the derivatives by central divided differences, yielding the DDF that is a class of derivative-free filters with $2d+1$ points ($\langle 0 \rangle$ and $\langle u_1 \rangle$) [10]. The DDF1 is based on the first-order Taylor polynomial approximation. In specific, it gives (see [10, eq. (24) and (26)])

$$\begin{aligned} I_{\mathrm{DDF1}}[\tilde{F}] &= \tilde{F}(0) = I_{\mathrm{EKF}}[\tilde{F}] \\ J_{\mathrm{DDF1}}[\tilde{F}_1, \tilde{F}_2] &= \frac{1}{4u_1^2} \sum_{i=1}^{d} \left(\tilde{F}_1(u_1 e_i) - \tilde{F}_1(-u_1 e_i)\right) \\ &\qquad \times \left(\tilde{F}_2(u_1 e_i) - \tilde{F}_2(-u_1 e_i)\right)^T \end{aligned} \tag{25}$$

where u_1 is the same as in (15). I_{DDF1} has precision 1; J_{DDF1} has precision 3. Note that J_{DDF1} is exact for some monomial

of degree 4 such as x_1^4 with $\tilde{F}_1(x) = x_1$ and $\tilde{F}_2(x) = x_1^3$ and all monomials odd in one or more coordinates. Based on the second-order Taylor polynomial approximation, the DDF2 is (see [10, eq. (30) and (40)])

$$\begin{aligned} I_{\mathrm{DDF2}}[\tilde{F}] &= \tilde{F}_1(0) + \frac{1}{2u_1^2} \\ &\quad \times \sum_{i=1}^{d} \left(\tilde{F}(u_1 e_i) + \tilde{F}(-u_1 e_i) - 2\tilde{F}(0) \right) \\ &= I_{\mathrm{UKF3}}[\tilde{F}] \\ J_{\mathrm{DDF2}}[\tilde{F}_1, \tilde{F}_2] &= \frac{1}{4u_1^2} \sum_{i=1}^{d} \left(\tilde{F}_1(u_1 e_i) - \tilde{F}_1(-u_1 e_i) \right) \\ &\quad \times \left(\tilde{F}_2(u_1 e_i) - \tilde{F}_2(-u_1 e_i) \right)^T \\ &\quad + \frac{1}{2u_1^4} \sum_{i=1}^{d} H_{i,i}^{(1)} \left(H_{i,i}^{(2)} \right)^T \\ &= J_{\mathrm{DDF1}}[\tilde{F}_1, \tilde{F}_2] \\ &\quad + \frac{1}{2u_1^4} \sum_{i=1}^{d} H_{i,i}^{(1)} \left(H_{i,i}^{(2)} \right)^T \end{aligned} \tag{26}$$

where $H_{i,i}^{(k)} = \tilde{F}_k(u_1 e_i) + \tilde{F}_k(-u_1 e_i) - 2\tilde{F}_k(0)$. I_{DDF2} has precision 3. Following the same line of (24), we can verify that J_{DDF2} is exact for monomials up to degree 3 but not for $x_1^2 x_2^2$ with $\tilde{F}_1(x) = \tilde{F}_2(x) = x_1 x_2$, so J_{DDF2} has precision 3. However, it should be noted that J_{DDF2} is indeed exact for all monomials of degree 4 but $x_1^2 x_2^2$ with $\tilde{F}_1(x) = \tilde{F}_2(x) = x_1 x_2$, and all monomials odd in one or more coordinates.

Almost at the same time, Ito *et al.* [7] proposed the CDF from the viewpoint of numerical integration. The CDF1 is based on the incomplete quadratic interpolatory polynomial that excludes monomials $x_i x_j$ with $i \neq j$. It can be proved that the CDF1 is essentially identical to the DDF2 [20]. On the other hand, the CDF2 is based on the quadratic interpolatory polynomial with $(d^2 + 3d + 2)/2$ points ($\langle 0 \rangle$, $\langle u_1 \rangle$ and $\langle u_1, u_1 \rangle^*$). In contrast to $\langle \cdot \rangle$, here $\langle \cdot \rangle^*$ denotes the set generated only by permutations. For example, if $x = [1, 0]^T$ then $\langle 1 \rangle^* = \{[1, 0]^T, [0, 1]^T\}$. Specifically (see [7, eq. (3.6) and (3.7)])

$$\begin{aligned} I_{\mathrm{CDF2}}[\tilde{F}] &= \tilde{F}_1(0) + \frac{1}{2u_1^2} \\ &\quad \times \left(\tilde{F}(u_1 e_i) + \tilde{F}(-u_1 e_i) - 2\tilde{F}(0) \right) \\ &= I_{\mathrm{CDF1}}[\tilde{F}] = I_{\mathrm{DDF2}}[\tilde{F}] = I_{\mathrm{UKF3}}[\tilde{F}] \\ J_{\mathrm{CDF2}}[\tilde{F}_1, \tilde{F}_2] &= \frac{1}{4u_1^2} \sum_{i=1}^{d} \left(\tilde{F}_1(u_1 e_i) - \tilde{F}_1(-u_1 e_i) \right) \\ &\quad \times \left(\tilde{F}_2(u_1 e_i) - \tilde{F}_2(-u_1 e_i) \right)^T \\ &\quad + \frac{1}{2u_1^4} \sum_{i=1}^{d} H_{i,i}^{(1)} \left(H_{i,i}^{(2)} \right)^T \\ &\quad + \frac{1}{u_1^4} \sum_{i,j=1, i<j}^{d} H_{i,j}^{(1)} \left(H_{i,j}^{(2)} \right)^T \\ &= J_{\mathrm{CDF2}}[\tilde{F}_1, \tilde{F}_2] \\ &\quad + \frac{1}{u_1^4} \sum_{i,j=1, i<j}^{d} H_{i,j}^{(1)} \left(H_{i,j}^{(2)} \right)^T \end{aligned} \tag{27}$$

where $H_{i,j}^{(k)} = \tilde{F}_k(u_1 e_i + u_1 e_j) - \tilde{F}_k(u_1 e_i) - \tilde{F}_k(u_1 e_j) + \tilde{F}_k(0)$. I_{CDF2} has precision 3 and J_{CDF2} has precision 5, which can be verified following the same line of (24).

IV. Comparative Study: Analysis and Simulation

A. Accuracy, Efficiency, and Stability Factor

The accuracy of a filter in terms of precision is summarized in the second column of Table I. The precision of the GHF depends on what number of points is used for the univariate integral. A list in the ascending order is

$$\begin{aligned} &\text{EKF, DDF1} < \text{2-point GHF, UKF3, DDF2/CDF1} \\ &\qquad < \text{CDF2} < \text{3-point GHF and UKF5.} \end{aligned} \tag{28}$$

Unsurprisingly, the overall reported results in the literature [1]–[3], [7], [10], [19], [21] accord to the list.

As compared with rules exact for monomials, methods of approximation make use of the fact that the covariance integrals have the special structure that the integrands are exclusively products of two functions (see (22)) and thus choose to address the multipliers separately. In contrast, rules exact for monomials treat any integral in an unaltered manner whatever the integrand is (see (10)). As a result, although of the same precision, the DDF2/CDF1 theoretically tends to outperform the UKF3 because J_{DDF2} is exact for all monomials of degree 4 but $x_1^2 x_2^2$ with $\tilde{F}_1(x) = \tilde{F}_2(x) = x_1 x_2$. This fact essentially explains why the CDF1 (or equivalently the DDF2) was reported to be superior to the UKF3 in [7, Example 8.2]. On the other hand, the UKF3 theoretically tends to outperform the 2-point GHF in that the UKF3 is exact for a monomial of degree 4 with the free parameter $u_1 = \sqrt{3}$; the DDF1 theoretically tends to outperform the EKF because J_{DDF1} is exact for some monomials of degree 4, such as x_1^4 with $\tilde{F}_1(x) = x_1$ and $\tilde{F}_2(x) = x_1^3$ while J_{EKF} is not. According to these analyses, a finer list is obtained as

$$\begin{aligned} &\text{EKF} < \text{DDF1} < \text{2-point GHF} < \text{UKF3} < \text{DDF2/ CDF1} \\ &\qquad < \text{CDF2} < \text{3-point GHF and UKF5.} \end{aligned} \tag{29}$$

It should be reminded that this is a theoretical conclusion and thus can only be served as a general guide. For a specific problem, the borderlines in between may become weakened (see the numerical examples in the sequel).

For efficiency comparison in the third column, we prefer to use the number of function evaluations as the criterion. An exception is the EKF that employs the derivative information instead of function values in approximating covariance integrals (see (23)). Its computational cost is comparable to that of the UKF3 because it has to calculate the Jacobian matrix [21]. It appears from Table I that the GHF is the most demanding filter in computation and the DDF1 is the least.

TABLE I
COMPARISONS OF APPROXIMATE GAUSSIAN FILTERS

Filter	Accuracy* (Mean, Covariance)	Efficiency (Number of Function Evaluations)	Stability Factor
m-point GHF (9)	$2m$-1	m^d	1
UKF3 (13)	3	$2d+1$	$\begin{cases} 1 & (d \le 3) \\ \frac{2d}{3}-1 & (d>3) \end{cases}$
UKF5 (14)	5	$2d^2+1$	$\begin{cases} 1 & (d \le 4) \\ 1+\frac{2d^2-8d}{9} & (d>4) \end{cases}$
EKF (21) and (23)	1, 3	1, —	1, —
DDF1 (25)	1, 3	1, $2d+1$	1, —
DDF2/CDF1 (26)	3, 3	$2d+1$, $2d+1$	$\begin{cases} 1 & (d \le 3) \\ \frac{2d}{3}-1 & (d>3) \end{cases}$, —
CDF2 (27)	3, 5	$2d+1$, $\left(d^2+3d+2\right)/2$	$\begin{cases} 1 & (d \le 3) \\ \frac{2d}{3}-1 & (d>3) \end{cases}$, —

* For filters based on the product rule and rules exact for monomials, the integration rules calculating mean and covariance are of the same precision; for filters based on methods of approximation, the precisions for mean and covariance are different and thus specified respectively. So are the efficiency and stability factor.

Another important issue that arises when using an integration rule is its stability[6] factor. Generally speaking from the perspective of integration, a rule with all of its points inside the integration region and all of its weights positive is more desirable and will have a small error than a formula for which one or both of these properties is lacking [33], [34]. For integrals of interest (6) and (22) in this paper, the first property is guaranteed. However, we are not so lucky with the second property of positive weights. For example, with the UKF3, the weight $w_0 = 1 - d/u_1^2$ in (13) will be negative when $d > u_1^2 = 3$. Integration rules with both positive and negative weights are particularly undesirable if the magnitude of the weights are large compared with $I(1)$. The reason is that a large amount of roundoff error is introduced if

$$\sum_{i=0}^{m} |w_i| \gg \sum_{i=0}^{m} w_i = I(1) = 1. \tag{30}$$

Therefore, a standard measure of the stability factor of an integration rule is the sum of the absolute values of the weights $\sum_{i=0}^{m} |w_i|/I(1) = \sum_{i=0}^{m} |w_i|$ [33], [37], [39]. A completely stable rule has $\sum_{i=0}^{m} |w_i| = 1$, but so far there is no known general method for constructing stable but efficient rules. Table I lists the stability factors in the last column. Note that the stability factor cannot be used to investigate the numerical stability of covariance calculations for the methods of approximation-based filters because it is defined to be used in the case that has only one integrand. Fortunately, the forms of J_{EKF} in (23), J_{DDF1} in (25), J_{DDF2} in (26), and J_{CDF2} in (27) assure in nature the positive definiteness of the resulting covariance. Of all Gaussian filters, the GHF is the only one that is completely stable but, as mentioned above, its computational cost increases exponentially with dimension. The EKF is numerically stable in the mean but we have to derive and compute the Jacobian matrix. The UKF, the DDF, and the CDF become numerically unstable when the dimension is larger than 3 or 4. The unstability might lead to the problem of yielding a nonpositive, semidefinite covariance in implementation, which is fatal for any Gaussian filter. Although the numerical unstability does not necessary incur failure of a filter, we have to take cautions when applying it to moderately large dimension systems. With the UKF, Julier *et al.* [18] proposed to use a modified form that evaluates the covariance about the projected mean. An alternative approach is to find the "closest" positive definite covariance matrix [40, Sec. 6.3]. Recently, the UKF has found a number of applications in practice [1]–[3], [7], [10], [19], [21], [44]–[46], in which the target systems are mostly of lower dimension. In view of the numerical unstability, the UKF is supposed to encounter troubles when being used to high-dimensional systems.

[6]The stability in the paper relates to the roundoff error [33, pp. 208–217] for integration calculation. It has little to do with the stability of a filter in the sense of mean-square error.

B. Numerical Examples

We next examine each Gaussian filter using three representative examples, viz. the univariate nonstationary growth model, the ballistic target reentry and bearing only tracking, all of which have significant nonlinearity and have been extensively investigated in the literature [7], [8], [18], [47]–[50]. We also presented the Cramer–Rao lower bound (CRLB) [26], [51] for the reference of performance comparison. The theoretical CRLB provides the best achievable error performance for any filter, which admits us to quantify how much scope is left to improve a filter. In specific, the covariance matrix of the estimate has a lower bound (referred to as the CRLB) as follows:

$$E\left\{(x_n - x_{n|n})^T (x_n - x_{n|n})\right\} \ge \mathrm{CRLB}(n). \tag{31}$$

1) Univariate Nonstationary Growth Model: This model is very popular in econometrics. It is formulated as

$$x_n = 0.5x_{n-1} + 25\frac{x_{n-1}}{1+x_{n-1}^2} + 8\cos(1.2(n-1)) + w_{n-1}$$
$$y_n = \frac{x_n^2}{20} + v_n, \quad n = 1, \ldots, N \tag{32}$$

where the process noise $w_{n-1} \sim \mathcal{N}(0,1)$ and the measurement noise $v_n \sim \mathcal{N}(0,1)$. This model has bimodality in nature, depending on the sign of observations. The reference data were generated using $x_0 = 0.1$ and $N = 500$. The CRLB was approximated by Monte Carlo averaging across 100 reference trajectories.

We have something to add before proceeding. Because the system (32) is of dimension $d = 1$, it is straightforward to prove that the UKF3, UKF5, and 3-point GHF are analytically identical to each other (see (13), (14), and (9)) and the CDF2 is analytically identical to the DDF2 as well (see (26) and (27)). Thus, for this system, the list in (29) becomes

$$\text{EKF} < \text{DDF1} < \text{2-point GHF} < \underline{\text{UKF3}} < \underline{\text{DDF2/CDF1}} = \underline{\text{CDF2}} < \underline{\underline{\text{3-point GHF and UKF5}}} \tag{33}$$

from which we can predict that the UKF3/UKF5/3-point GHF will be "identical" to the CDF1/CDF2/DDF2.

For each filter, the initial conditions were $x_{0|0} = 0$, $P_{0|0} = 1$. The performance was compared using the mean-square error (MSE) defined by

$$\text{MSE} = \frac{1}{N}\sum_{n=1}^{N}(x_n - x_{n|n})^2. \tag{34}$$

The MSEs of 50 Monte Carlo runs are plotted in Fig. 2, as well as the MSE derived from the CRLB by averaging across N time instants, i.e.,

$$\text{CRLB-MSE} = \frac{1}{N}\sum_{n=1}^{N}\text{CRLB}(n). \tag{35}$$

In Fig. 2, The CRLB-MSE is by far smaller than that of any filter, which indicates all of the above Gaussian filters are not good enough to address this system. However, this does not hinder the comparison among Gaussian filters.

The mean and standard variance of MSEs are plotted in Fig. 3. The UKF3/UKF5/3-point GHF is not distinguishable from the CDF2/DDF2, which shows that the above prediction is right. The EKF is the largest, and the DDF1 is the second largest in MSE. The 2-point GHF comes as the third, but its standard variance of MSEs is slightly larger than that of the DDF1. We see that the observations agree rather well with the analytical conclusion in (33).

2) Ballistic Target Re-entry: The aim is to estimate the position $x_1(t)(ft)$, velocity $x_2(t)(ft/s)$, and constant ballistic coefficient $x_3(t)(1/s)$ of a body as it re-enters the atmosphere at a very high altitude with high speed. The motion is determined by altitude- and velocity-related drag terms, and the body is constrained to fall vertically. The position of the body is measured at

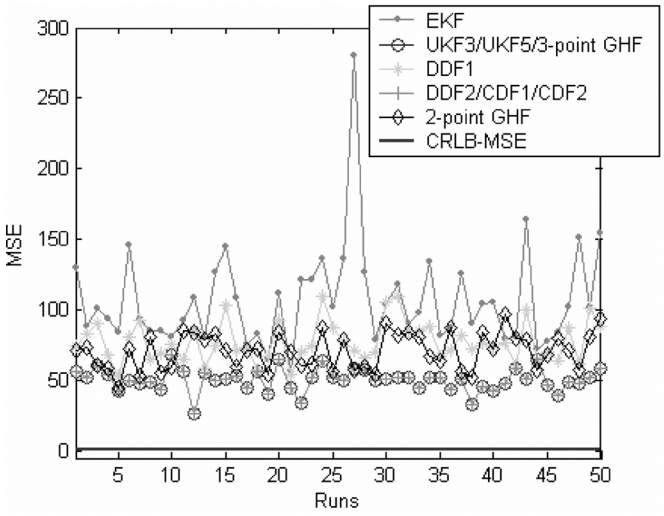


Fig. 2. MSEs of each filter across 50 random runs and the MSE derived from the CRLB.

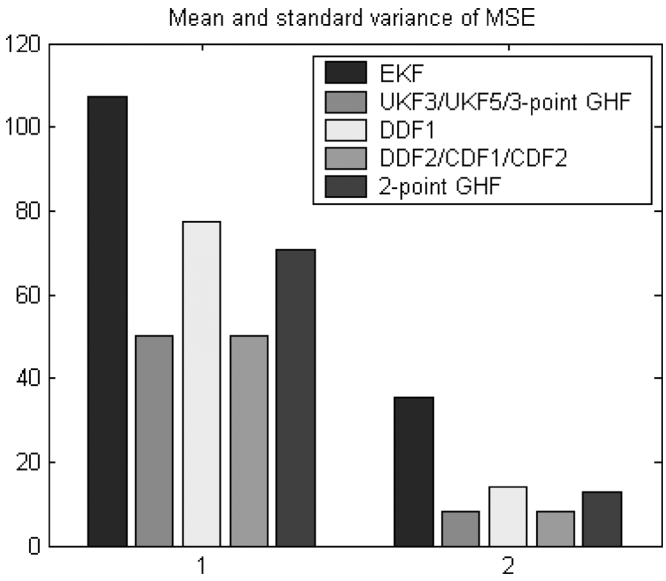


Fig. 3. Mean and standard variance of MSEs across 50 random runs.

discrete time instants using a radar capable of measuring range contaminated by Gaussian noise. The radar locates at an altitude $H(ft)$, and the horizontal distance between the body and the radar is $M(ft)$.

The continuous time dynamics of the system are

$$\begin{aligned}\dot{x}_1(t) &= -x_2(t) + w_1(t)\\ \dot{x}_2(t) &= -e^{-\gamma x_1(t)}x_2^2(t)x_3(t) + w_2(t)\\ \dot{x}_3(t) &= w_3(t)\end{aligned} \tag{36}$$

where $[w_1(t), w_2(t), w_3(t)]^T \sim N(0, Q)$ and the constant γ relates the air density with altitude. The range is

$$y(t) = \sqrt{M^2 + (x_1(t) - H)^2} + r(t) \tag{37}$$

where $r(t) \sim \mathcal{N}(0, R)$. The measurements are made at the frequency of 1 Hz.

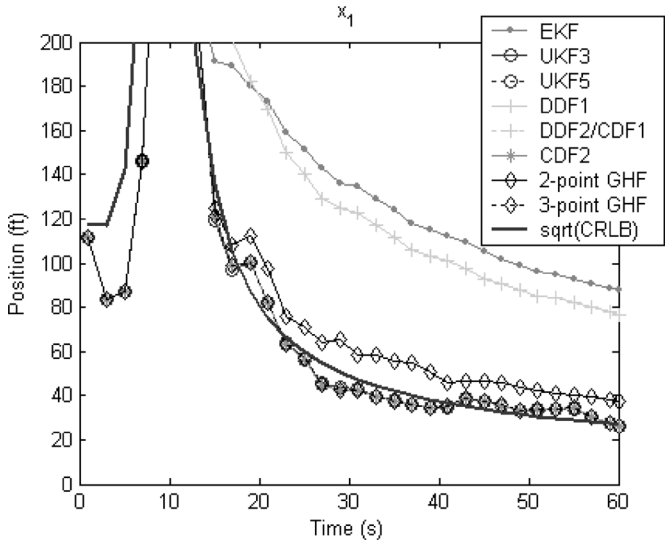


Fig. 4. Averaged absolute position error (UKF3, UKF5, DDF2/CDF1, CDF2, and 3-point GHF overlap each other).

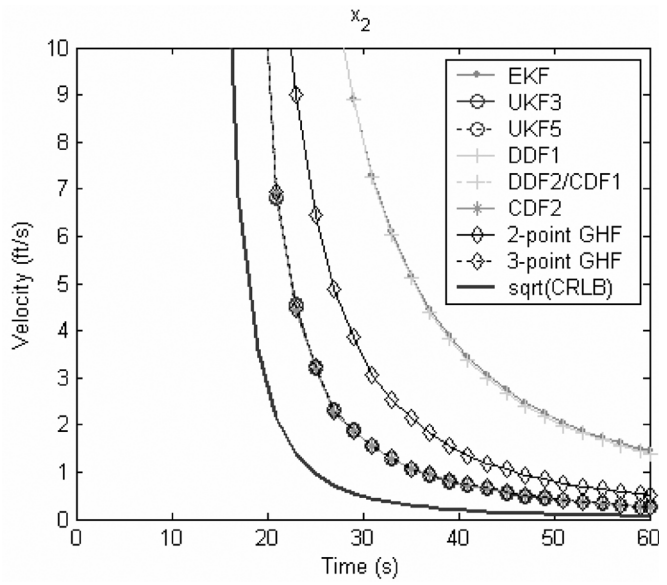


Fig. 5. Averaged absolute velocity error (UKF3, UKF5, DDF2/CDF1, CDF2, and 3-point GHF overlap each other).

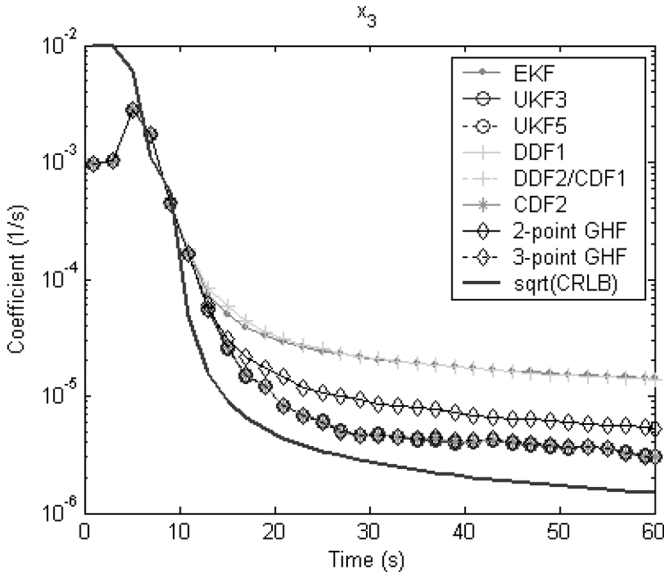


Fig. 6. Averaged absolute coefficient error (UKF3, UKF5, DDF2/CDF1, CDF2, and 3-point GHF overlap each other).

In the previous literature, a fourth-order Runge–Kutta scheme with 64 steps between each observation was employed to integrate (36) in order to deal with the significant nonlinearity of the system [7], [18], [50]. For the EKF, it was necessary to recalculate the Jacobian 64 times between each update; for the other derivative-free filters, the predication at each point was calculated using the small steps, and it was only necessary to calculate the mean and covariance just before each measurement arrived.

The following data were used: the system parameters were chosen as $\gamma = 5 \times 10^{-5}$, $M = 10^5$, $H = 10^5$, $Q = 0$, and $R = 10^4$. The initial true state of the system was $x_0 = [3 \times 10^5, 2 \times 10^4, 10^{-3}]^T$, and the initial estimate is $x_{0|0} = [3 \times 10^5, 2 \times 10^4, 3 \times 10^{-5}]^T$ with covariance

$$P_{0|0} = \begin{bmatrix} 10^6 & 0 & 0 \\ 0 & 4 \times 10^6 & 0 \\ 0 & 0 & 10^{-4} \end{bmatrix}.$$

For this example, we used the averaged absolute error to compare each filter. The averaged absolute error of the kth component of the state at time instant n is defined by

$$\varepsilon_{(k)}(n) = \frac{1}{L} \sum_{l=1}^{L} \left| x_{(k),n}^{(l)} - x_{(k),n|n}^{(l)} \right| \tag{38}$$

where L is the number of Monte Carlo runs. In the absence of process noise, the CRLB recursion for the system (36) and (37) is identical to the covariance matrix propagation of the EKF, where the Jacobians are evaluated at the true state.

In Figs. 4–6, we show the averaged absolute error for each component committed by each filter across a simulation consisting of 50 Monte Carlo runs, as well as the $\sqrt{\text{CRLB}}$. These figures show that the EKF and DDF1 are the least accurate filters with the latter being slightly better than the former. The 2-point GHF has a noticeable improvement over the EKF and DDF1. Other filters, namely, the UKF3, UKF5, DDF2/CDF1, CDF2, and 3-point GHF, come next as the best, approaching optimality in position estimation. There is no discernible difference in accuracy from each other. Recall the fact that although of precision 3, the UKF3, DDF2/CDF1 and CDF2 are indeed exact for the fourth-order principle moment. The result indicates that a rule's inexactness for the fourth-order cross moments has little influence on performance for this example. On the other hand, the improvement of the 2-point GHF over the EKF and DDF1 implies the significance of the second-order moment; the further improvement of the UKF3, UKF5, DDF2/CDF1, CDF2, and 3-point GHF over the 2-point GHF implies the significance of the fourth-order principle moment. These insights will be helpful to the choice of a filter in practice.

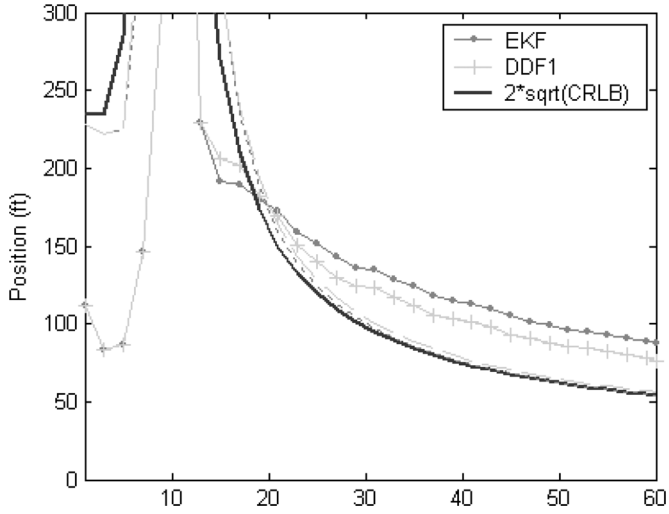


Fig. 7. Averaged absolute position error with the associated 2σ-deviation bound (dotted line for the EKF and dashed line for the DDF1).

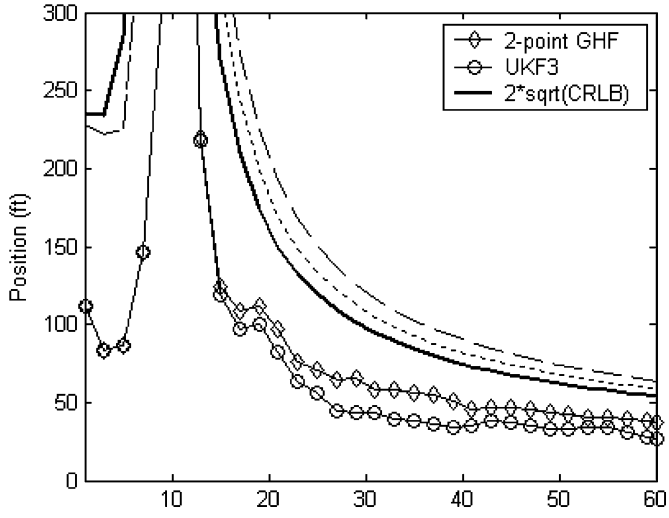


Fig. 8. Averaged absolute position error with the associated 2-standard deviation bound (dotted line for the 2-point GHF and dashed line for the UKF3).

Fig. 7 plots the errors of the EKF and DDF1 in position estimates, the associated estimates of the 2σ-deviation bounds and $2\times\sqrt{\text{CRLB}}$. Since the UKF3, UKF5, DDF2/CDF1, CDF2, and 3-point GHF are quite similar in accuracy, Fig. 8 only plots the 2-point GHF and UKF3 for clarity. The 2σ-deviation bounds are given by twice the averaged square root of the diagonals of the covariance matrix. The estimate error of a consistent filter should lie below its corresponding bound with a possibility of 0.95. Obviously, the EKF and DDF1 do not produce consistent estimates after 20 s or so (see Fig. 7) and are too optimistic with the estimate. The 2-point GHF and UKF3 stay well below their associated bounds for the whole time span (see Fig. 8).

3) Bearing Only Tracking: The target moves within the $s-t$ plane according to the standard second-order model

$$x_n = \Phi x_{n-1} + \Gamma w_{n-1}, \quad n = 1, \ldots, N \tag{39}$$

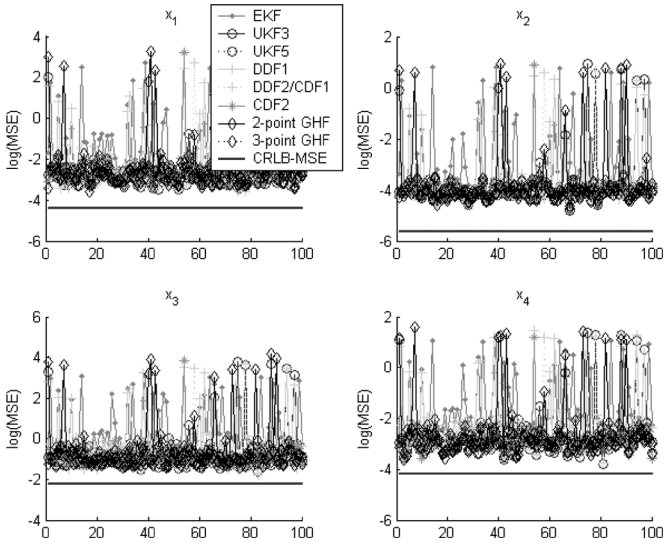


Fig. 9. MSEs of each filter across 100 random runs and the MSE derived from the CRLB.

where $x_n = [s, \dot{s}, t, \dot{t}]_n^T$, $w_n = [w_s, w_t]_n^T$

$$\Phi = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} 0.5 & 0 \\ 1 & 0 \\ 0 & 0.5 \\ 0 & 1 \end{bmatrix}.$$

Here, s and t denote Cartesian coordinates of the moving target. The system noise $w_n \sim \mathcal{N}(0, QI_2)$, where I_2 is the 2×2 identity matrix. A fixed observer at the origin of the plane takes noisy measurements of the target bearing

$$y_n = \arctan(t_n/s_n) + v_n \tag{40}$$

where the measurement noise $v_n \sim \mathcal{N}(0, R)$. Obviously, the state is unobservable for certain target-observer geometries, making the posterior probability density highly non-Gaussian. A Gaussian filter performs satisfactorily only when the Gaussian assumption is valid. If not, a Gaussian filter would risk divergence.

The reference data were generated using $Q = 0.001^2$, $R = 0.005^2$. The initial true state of the system was $x_0 = [-0.05, 0.001, 0.7, -0.055]^T$, and the initial estimate was $x_{0|0} = [-0.04, 0, 0.6, -0.05]^T$ with covariance $P_{0|0} = \text{diag}([0.1^2, 0.005^2, 0.1^2, 0.01^2]^T)$. Realistic amounts of process noise make a very small impact on the CRLB for bearing only tracking [51], so the CRLB carried out for this example assumed zero process noise.

The MSEs of 100 Monte Carlo runs for each component are plotted in Fig. 9, as well as the MSE derived from the CRLB by averaging across N time instants as shown in (35). It appears that divergence occurs for most of filters. In such a case, we turn to use for comparison the number of runs of divergence, instead of the MSE. Table II provides the number of runs of divergence for each filter out of 500 Monte Carlo implementations. The result agrees very well with the analytical ranking in

TABLE II
NUMBER OF RUNS OF DIVERGENCE FOR EACH FILTER

	EKF	UKF3	UKF5	DDF1	DDF2	CDF2	GHF2	GHF3
s	118	26	1	71	25	0	42	0
$\dot{s}$	147	29	1	80	25	1	49	0
t	130	31	1	79	26	2	49	0
$\dot{t}$	125	31	1	77	26	2	48	0

(29). The 3-point GHF is the only one that did not diverge in our simulation.

V. DISCUSSIONS AND CONCLUSION

This paper reviews the Gaussian filters from the perspective of numerical integration. Specifically, we present in a unified numerical-integration framework the derivation of a number of approximate Gaussian filters. It shows that all Gaussian filters are approximations of the general Gaussian filter by using a specific numerical integration method of some kind or another, such as the Gauss–Hermite product rule, rules exact for monomials and methods of approximation. This perspective provides a well-founded understanding of all the existing Gaussian filters with respect to accuracy, efficiency, and stability factor. The analytical findings are tabulated, from which a ranking of accuracy of various Gaussian filters is derived. The numerical results agree nicely with the analytical ranking list. We believe that this perspective will facilitate selection of Gaussian filters in practice and hopefully be useful to design more efficient and stable filters by employing better numerical integration methods.

Given a nonlinear system, the ideal scenario of selecting a suitable Gaussian filter could be the following.

1) Confirm the validity of the Gaussian assumption. This should be done with a CRLB-like bound that tells the best achievable performance for any Gaussian filter using the linear update rule. The development of the bound remains an open problem. Regardless of its Bayesian updating, the GPF equipped with sufficient samples can be expected to play this role.
2) Select a candidate Gaussian filter according to Table I. It depends on which aspect or aspects most matter. For example, if the accuracy alone is the most concerned, it will be advisable to select the 3-point GHF or the UKF5.
3) Apply the candidate to the target system to see whether it is qualified. If not, there will be no need to carry out other Gaussian filters, and we have to design new algorithms of higher precision or directly turn to use sampling-based filters, e.g., the GPF or the general particle filter.

The work in this paper is an essential step to put the above scenario into reality.

ACKNOWLEDGMENT

The authors would like to thank the coordinating editor and anonymous reviewers for their insightful comments and constructive suggestions that considerably improved this paper, especially one reviewer for his very detailed annotations.

REFERENCES

[1] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," *J. Guid. Control Dyn.*, vol. 26, no. 4, pp. 536–542, 2003.
[2] A. Farina, B. Ristic, and D. Benvenuti, "Tracking a ballistic target: comparison of several nonlinear filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 854–867, Jul. 2002.
[3] B. Ristic, A. Farina, D. Benvenuti, and M. S. Arulampalam, "Performance bounds and comparison of nonlinear filters for tracking a ballistic object on re-entry," *Proc. Inst. Elect. Eng.—Radar Sonar Navigation*, vol. 150, no. 2, pp. 65–70, 2003.
[4] M. O'Connor, G. Elkaim, T. Bell, and B. Parkinson, "Automatic steering of farm vehicles using GPS," in *Proc. 3rd Int. Conf. Precision Agriculture*, 1996, pp. 767–778.
[5] N. D. Tillet, "Automatic guidance for agricultural field machines: A review," *J. Agricult. Eng. Res.*, vol. 50, pp. 167–187, 1991.
[6] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Autom. Control*, vol. AC-9, no. 4, pp. 333–339, Oct. 1964.
[7] K. Ito and K. Q. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.
[8] J. H. Kotecha and P. A. Djuric, "Gaussian particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.
[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, vol. 82, pp. 34–45, 1960.
[10] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, 2000.
[11] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," *Proc. SPIE-Int. Soc. Opt. Eng.—Signal Processing, Sensor Fusion, Target Recognition VI*, vol. 3068, pp. 182–193, 1997.
[12] A. H. Jazwinski, *Stochastic Processing and Filtering Theory*. New York: Academic, 1970.
[13] H. W. Sorenson and A. R. Stubberud, "Nonlinear filtering by approximation of the a posterior density," *Int. J. Control*, vol. 8, pp. 33–51, 1968.
[14] S. C. Kramer and H. W. Sorenson, "Recursive Bayesian estimation using piece-wise constant approximations," *Automatica*, vol. 34, no. 6, pp. 789–801, 1988.
[15] S. F. Schmidt, "The Kalman filter—Its recognition and development for aerospace applications," *J. Guid. Control Dyn.*, vol. 4, no. 1, pp. 4–7, 1981.
[16] R. H. Battin, "Some funny things happened on the way to the moon," *J. Guid. Control Dyn.*, vol. 25, no. 1, pp. 1–7, 2002.
[17] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.
[18] S. Julier, J. Uhlmann, 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. 477–482, Mar. 2000.
[19] E. A. Wan and R. v. d. Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. New York: Wiley, 2001, ch. 7, pp. 221–280.
[20] R. v. d. Merwe, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," Ph.D. dissertation, Electrical and Computer Eng. Dept., Oregon Health & Science Univ., 2004.
[21] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
[22] 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.
[23] A. F. M. Smith, A. M. Skene, J. E. H. Shaw, J. C. Naylor, and M. Dransfield, "The implementation of the Bayesian paradigm," *Commun. Statistics—Theory Meth.*, vol. 14, pp. 1079–1102, 1985.
[24] S. R. McReynolds, "Multidimensional Hermite–Gaussian quadrature formulae and their application to nonlinear estimation," in *Proc. 6th Symp. Nonlinear Estimation*, 1975, pp. 188–191.
[25] Y. Wu, X. Hu, D. Hu, and M. Wu, "Comments on "Gaussian particle filtering"," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3350–3351, Aug. 2005.
[26] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer–Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[27] R. Mehra, "A comparison of several non-linear filters for re-entry vehicle tracking," *IEEE Trans. Autom. Control*, vol. AC-16, no. 4, pp. 307–319, 1971.
[28] T. Lefebvre, H. Bruyninckx, and J. D. Schutter, "Kalman filters for nonlinear systems: A comparison of performance," *Int. J. Control*, vol. 77, no. 7, pp. 639–653, 2004.
[29] F. L. Lewis, *Optimal Estimation*. New York: Wiley, 1986.
[30] J. E. H. Shaw, A. Genz, J. Monahan, M. J. Schervish, L. Wasserman, R. Wolfinger, M. Evans, and T. Swartz, "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems—Comments and rejoinders," *Stat. Sci.*, vol. 11, no. 1, pp. 54–64, 1996.
[31] M. Evans and T. Swartz, "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems," *Stat. Sci.*, vol. 10, no. 3, pp. 254–272, 1995.
[32] W. H. Press, *Numerical Recipes: The Art of Scientific Computing*. Cambridge, MA: Cambridge Univ. Press, 1989.
[33] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. New York: Academic, 1975.
[34] A. H. Stroud, *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[35] J. McNamee and F. Stenger, "Construction of fully symmetric numerical integration formulas," *Numer. Math.*, vol. 10, pp. 327–344, 1966.
[36] R. Cools and P. Rabinowitz, "Monomial cubature rules since Stroud—A compilation," *J. Comput. Appl. Math.*, vol. 48, no. 3, pp. 309–326, 1993.
[37] A. Genz and B. D. Keister, "Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight," *J. Comput. Appl. Math.*, vol. 71, no. 2, pp. 299–309, 1996.
[38] E. Novak, K. Ritter, R. Schmitt, and A. Steinbauer, "On an interpolatory method for high dimensional integration," *J. Comput. Appl. Math.*, vol. 112, no. 1–2, pp. 215–228, 1999.
[39] S. Capstick and B. D. Keister, "Multidimensional quadrature algorithms at higher degree and/or dimension," *J. Comput. Phys.*, vol. 123, no. 2, pp. 267–273, 1996.
[40] U. N. Lerner, "Hybrid Bayesian networks for reasoning about complex systems," Ph.D. dissertation, Computer Sci. Dept., Stanford Univ., Stanford, CA, 2002.
[41] S. J. Julier and J. K. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations," in *Proc. Amer. Control Conf.*, 2002, pp. 887–892.
[42] S. Julier, "The spherical simplex unscented transformation," in *Proc. Amer. Control Conf.*, 2003, vol. 3, pp. 2430–2434.
[43] Private communication, A. Genz, Dept. of Mathematics, Washington State Univ., 2004, 9.
[44] S. J. Julier, "Comprehensive process models for high-speed navigation," Ph.D. dissertation, Eng. Sci. Dept., Univ. of Oxford, Oxford, U.K., 1996.
[45] S. Clark, "Autonomous land vehicle navigation using millimeter wave radar," Ph.D. dissertation, Mechanical Mechatronic Eng. Dept., Univ. of Sydney, Sydney, Australia, 1999.
[46] W. Li, H. Leung, and Y. Zhou, "Space-time registration of radar and ESM using unscented Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 824–836, Jul. 2004.
[47] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time-series," *J. Amer. Stat. Assoc.*, vol. 82, no. 400, pp. 1032–1063, 1987.
[48] E. R. Beadle and P. M. Djuric, "A fast weighted Bayesian bootstrap filter for nonlinear model state estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 1, pp. 338–343, Jan. 1997.
[49] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng.—F*, vol. 140, no. 2, pp. 107–113, 1993.
[50] M. Athans, R. P. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous time nonlinear systems from discrete noisy measurements," *IEEE Trans. Autom. Control*, vol. 13, no. 5, pp. 504–514, Oct. 1968.
[51] B. Ristic and S. Arulampalm, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA: Artech House, 2004.

Yuanxin Wu was born in Jinan, China, in 1976. He received the B.Sc. and Ph.D. degrees both with honors in navigation, guidance, and control from the Department of Automatic Control, National University of Defense Technology, Changsha, China, in 1998 and 2005, respectively.

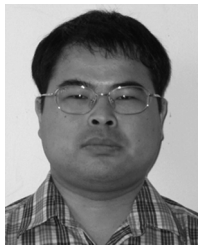
His current research interests include navigation system, general estimation theory, statistical signal processing, and application Clifford algebra in engineering.

Mr. Wu was awarded the Guanghua Scholarship twice in 1998 and 2003 and was the only student in the department to receive the University Ph.D. Innovation Fellowship in 2001.

Dewen Hu (M'03–SM'06) was born in Hunan, China, in 1963. He received the B.Sc. and M.Sc. degrees in electrical engineering from Xi'an Jiaotong University, China, in 1983 and 1986, respectively, and the Ph.D. degree in pattern analysis and intelligent systems from the University of Defense Technology, Changsha, China, in 1999.

From 1986 to 1995, he was with the National University of Defense Technology. From October 1995 to October 1996, he was a Visiting Scholar with the University of Sheffield, U.K. He is currently a Professor of the Department of Automatic Control with the National University of Defense Technology. His research interests include image processing, system identification and control, neural networks, and cognitive science. He has over 80 papers and three monographs in the areas of his interests.

Dr. Hu is the joint recipient of more than a dozen academic prizes for his research on neurocontrol, intelligent robot, and image processing.



Meiping Wu was born in Fujian, China, in 1971. He received the B.Sc. and M.Sc. degrees in aviation mechanics and the Ph.D. degree in navigation, guidance, and control from the Department of Automatic Control, National University of Defense Technology, Changsha, China, in 1993, 1996, and 2000, respectively.

Currently, he is an Associate Professor and the Associate Director of the Department of Automatic Control with the National University of Defense Technology. He is an expert of the 863 High Technology Project. His scientific interests include aircraft navigation, guidance, and control.

Xiaoping Hu was born in Sichuan, China, in 1960. He received the B.Sc. and M.Sc. degrees in automatic control systems and aerocraft designing from the Department of Automatic Control, National University of Defense Technology, Changsha, China, in 1982 and 1985, respectively.

Currently, he is a Professor and the Dean of College of Mechatronics and Automation with the National University of Defense Technology. His scientific interests include inertial and satellite navigation, aerocraft guidance, and control.