SIAM/ASA J. Uncertainty Quantification
Vol. 2, pp. 276–304

Gauss von Mises Distribution for Improved Uncertainty Realism in Space Situational Awareness*

Joshua T. Horwood† and Aubrey B. Poore†

Abstract. In order to provide a more statistically rigorous treatment of uncertainty in the space surveillance tracking environment, a new class of multivariate probability density functions is proposed called the Gauss von Mises (GVM) family of distributions. The distinguishing feature of the GVM distribution is its definition on a cylindrical manifold, the underlying state space in which systems of orbital element coordinates are more accurately defined. When specialized to GVM distributions, the prediction step of the general Bayesian nonlinear filter is shown to be tractable and viable, thereby providing a novel means to propagate space object orbital uncertainty under nonlinear perturbed two-body dynamics. Results demonstrate that uncertainty propagation with the new GVM distribution can be achieved at the same cost as the traditional unscented Kalman filter and can maintain "uncertainty realism" for up to eight times as long.

Key words. covariance realism, uncertainty realism, uncertainty propagation, nonlinear filtering, directional statistics, space surveillance

AMS subject classifications. 62H11, 62H12, 70F05, 70M20

DOI. 10.1137/130917296

1. Introduction.

Space situational awareness (SSA) is the comprehensive knowledge of the near-Earth space environment accomplished through the tracking and identification of orbiting space objects needed to protect space assets and maintain awareness of potentially adversarial space deployments. The *proper characterization of uncertainty* in the orbital state of a space object is a common requirement to many SSA functions, including tracking and data association, conjunction analysis and probability of collision, sensor resource management, and anomaly detection. While some tracking environments, such as air and missile defense, make extensive use of Gaussian and local linearity assumptions within uncertainty quantification algorithms, space surveillance is inherently different due to long time gaps between updates, high misdetection rates, nonlinear and nonconservative dynamics, and non-Gaussian phenomena. The latter implies that *covariance realism* or *covariance consistency* [3], the proper characterization of an orbit's mean state and covariance, is not always sufficient. SSA also needs a generalization of this concept called *uncertainty realism*, the proper characterization of the state and covariance (as required for covariance realism) in addition to all higher-order cumulants. In other words, uncertainty realism requires a proper characterization of a space object's full state *probability density function* (PDF) in order to faithfully represent the statistical errors.

*Received by the editors April 16, 2013; accepted for publication (in revised form) May 13, 2014; published electronically July 16, 2014. This work was funded, in part, by a Phase II STTR (FA9550-12-C-0034) and a grant (FA9550-11-1-0248) from the Air Force Office of Scientific Research.
http://www.siam.org/journals/juq/2/91729.html

†Numerica Corporation, Fort Collins, CO 80528 (joshua.horwood@numerica.us, aubrey.poore@numerica.us).

In order to provide a more statistically rigorous treatment of uncertainty in the space surveillance tracking environment and to better support the aforementioned SSA functions, a new class of multivariate PDFs, called *Gauss von Mises* (GVM) distributions, are formulated to more faithfully characterize the uncertainty of a space object's orbital state and hence provide improved uncertainty realism. Using the new GVM distribution as input, extensions and improvements are possible to many key tracking algorithms, including the Bayesian nonlinear filter used for uncertainty propagation and data fusion, batch processing and orbit determination, and likelihood ratios and other scoring metrics used in data association.

It has been recognized in the space surveillance community that the orbital state uncertainty of a space object can be highly non-Gaussian, and statistically robust methods for treating such non-Gaussian uncertainties are sometimes required. Examples of estimation and filtering techniques beyond the traditional extended Kalman filter (EKF) [12] and unscented Kalman filter (UKF) [13], both of which make local linearity and Gaussian assumptions, include Gaussian sum (mixture) filters [2, 9, 10, 26], filters based on nonlinear propagation of uncertainty using Taylor series expansions of the solution flow [5, 20, 21], and particle filters [24]. A drawback of many existing methods for nonlinear filtering and uncertainty quantification, including those listed above, is the constraint that the state space be defined on an n-dimensional Cartesian space $\mathbb{R}^n$. Any statistically rigorous treatment of uncertainty must use PDFs defined on the underlying manifold on which the system state is defined. In the space surveillance tracking problem, the system state is often defined with respect to *orbital element coordinates* [19]. In these coordinates, five of the six elements are approximated as unbounded Cartesian coordinates on $\mathbb{R}^5$, while the sixth element is an angular coordinate defined on the circle $\mathbb{S}$ with the angles θ and $\theta + 2\pi k$ (where k is any integer) identified as equivalent (i.e., they describe the same location on the orbit). Thus, more rigorously, an orbital element state space is defined on the six-dimensional cylinder $\mathbb{R}^5 \times \mathbb{S}$. Indeed, the mistreatment of an angular coordinate as an unbounded Cartesian coordinate can lead to many unexpected software faults and other dire consequences, as described in section 2. A key innovation of the GVM distribution is its definition on a *cylindrical state space* $\mathbb{R}^n \times \mathbb{S}$ with the proper treatment of the angular coordinate within the general framework of directional statistics [17]; hence, the GVM distribution is robust for uncertainty quantification in orbital element space. The GVM distribution uses the von Mises distribution [16, 17], the analogy of a Gaussian distribution defined on a circle, to robustly describe uncertainty in the angular coordinate. Additionally, the GVM distribution contains a parameter set controlling the correlation between the angular and Cartesian variables as well as the higher-order cumulants, which gives the level sets of the GVM PDF a distinctive "banana" or "boomerang" shape. Such level sets in orbital state PDFs have been observed in previous work [9].

By providing a statistically robust treatment of the uncertainty in a space object's orbital element state by rigorously defining the uncertainty on a cylindrical manifold, the GVM distribution supports a suite of next-generation algorithms for uncertainty propagation, data association, space catalog maintenance, and other SSA functions. When adapted to state PDFs modeled by GVM distributions, the general Bayesian nonlinear filter is tractable. The prediction step of the resulting GVM filter is made derivative-free, like the UKF, by new quadrature rules for integrating a function multiplied by a GVM weight function, thereby extending the unscented transform. Moreover, prediction using the GVM filter requires the

propagation of the same number of sigma points (quadrature nodes) as the standard UKF. Thus, the GVM filter prediction step (uncertainty propagation) has the *same computational cost as the UKF*, and, as demonstrated in section 7, the former can maintain a *proper characterization of the uncertainty* for up to eight times as long as the latter. In the most exceptional cases when the actual state uncertainty deviates from a GVM distribution, a mixture version of the GVM filter can be formulated (using GVM distributions as the mixture components) to ensure proper uncertainty realism in analogy to the Gaussian sum (mixture) filter. A maximum a posteriori batch processing capability for *orbit determination* (track initiation) can also be formulated which generates a GVM PDF characterizing the initial orbital state and uncertainty from a sequence of input reports such as radar, electro-optical, or IR sensor observation data or even full track states. To support the *data fusion* problem of tracking, the correction step of the Bayesian nonlinear filter can also be specialized to GVM distributions, thereby enabling one to combine reports emanating from a common object to improve the state or understanding of that object. The filter correction step also furnishes a statistically rigorous *prediction error* which appears in the likelihood ratios for scoring the association of one report to another [22]. Thus, the new GVM filter can be used to support *multitarget tracking* within a general multiple hypothesis tracking framework [22, 23]. Additionally, the GVM distribution admits a *distance metric* which extends the classical Mahalanobis distance (χ^2 statistic) [15]. This new "Mahalanobis von Mises" metric provides a test for statistical significance and facilitates validation of the GVM filter.

For this initial paper on the new GVM distribution, the focus will be on its motivation, definition, and statistical properties followed by the development of the uncertainty propagation (filter prediction) algorithm used in the resulting GVM filter. It is a continuation of an earlier conference proceedings paper [11] that provided the motivation and definition of the GVM distribution. Future publications will develop the filter correction step and other extensions discussed above. The plan of the paper is as follows. Section 2 gives an overview of uncertainty characterization in SSA and discusses coordinate systems used to describe a space object's orbital state and the pitfalls of mistreating an angular coordinate as an unbounded Cartesian coordinate. Section 3 motivates and defines the GVM distribution. Section 4 lists important mathematical properties of the GVM distribution. Section 5 extends the methodology of classical Gauss–Hermite quadrature to enable the computation of the expected value of a nonlinear transformation of a GVM random variable, thereby providing a general framework for GVM quadrature. Such quadrature formulas play a key role in the derivation of the GVM filter prediction step. Section 6 develops the prediction step of the Bayesian nonlinear filter using the GVM distribution. Section 7 demonstrates proof-of-concept of the GVM uncertainty propagation algorithm. Finally, section 8 provides concluding remarks.

2. Uncertainty on manifolds.

A permeating theme throughout SSA is the achievement of the correct characterization and quantification of uncertainty, which in turn is necessary to support conjunction analysis, data association, anomaly detection, and sensor resource management. Such success can depend greatly on the choice of *coordinate system*. Under Gaussian assumptions, the coordinates used to represent the state space can impact how long one can propagate the uncertainty under a nonlinear dynamical system; a Gaussian random vector will not get mapped to a Gaussian under a nonlinear transformation. The

representation of a space object's kinematic state in *orbital element coordinates* [19], rather than Cartesian Earth-centered-inertial (ECI) position-velocity coordinates, is well suited to the space surveillance tracking problem since such coordinates "absorb" the most dominant term in the nonlinear gravitational potential (i.e., the $1/r$ term), leading to "more linear" propagations. Thus, these special coordinates can mitigate the departure from "Gaussianity" under the nonlinear propagation of an initial Gaussian state PDF with respect to orbital elements. Additional discussions are provided in subsection 2.1.

The application of orbital element coordinates within traditional sequential filtering methods, such as the EKF, UKF, and even the highest fidelity Gaussian sum filters (GSFs), has one major disadvantage: the mean anomaly (or mean longitude) angular coordinate describing the location along the orbit is incorrectly treated as an unbounded Cartesian coordinate. Some side effects and pitfalls of such mistreatment within the problems of averaging and fusing angular quantities are described in subsection 2.2. Ultimately, what is required to rectify these shortcomings is a statistically rigorous treatment of the uncertainty on the *underlying manifold* on which the system state is defined. The theory of directional statistics [17] provides one possible development path. Though the theory can treat uncertainty on very general manifolds (such as tori and hyperspheres) possessing multiple directional quantities, this work focuses on distributions defined on the circle $\mathbb{S}$ and the $n+1$-dimensional cylinder $\mathbb{R}^n \times \mathbb{S}$. The latter is the manifold on which the orbital element coordinates are more accurately defined. As a starting point, the von Mises PDF [16, 17] is reviewed in subsection 2.3 as one possible distribution to rigorously treat the uncertainty of a single angular variable defined on the circle. The von Mises distribution paves the way for the next section concerning the development of the GVM distribution used to represent uncertainty on a cylinder.

2.1. Orbital element coordinate systems. With respect to Cartesian ECI position-velocity coordinates $(\boldsymbol{r}, \dot{\boldsymbol{r}})$, the acceleration $\ddot{\boldsymbol{r}}$ of a space object (e.g., a satellite or debris) can be written in the form

$$\ddot{\boldsymbol{r}} = -\frac{\mu_\oplus}{r^3}\boldsymbol{r} + \boldsymbol{a}_{pert}(\boldsymbol{r}, \dot{\boldsymbol{r}}, t). \tag{2.1}$$

In this equation, $r = |\boldsymbol{r}|$, $\mu_\oplus = GM_\oplus$, where G is the gravitational constant and $M_\oplus$ is the mass of the Earth, and $\boldsymbol{a}_{pert}$ encapsulates all perturbing accelerations of the space object other than those due to the two-body point mass gravitational acceleration.

The *Keplerian orbital elements* $(a, e, i, \Omega, \omega, M)$ define a system of curvilinear coordinates with respect to six-dimensional position-velocity space and encompass the geometric and physical properties of unperturbed two-body dynamics (i.e., (2.1) with $\boldsymbol{a}_{pert} = \boldsymbol{0}$). As dictated by Kepler's first law, any (closed) orbit is necessarily elliptical; the semimajor axis a and the eccentricity e quantify the geometry of ellipse. The orientation of the ellipse relative to the equatorial plane is described by the three Euler angles (Ω, i, ω) in the zxz convention, namely, the argument of perigee ω, the inclination i, and the longitude of the ascending node Ω. Finally, the mean anomaly M can be viewed as describing the location of the orbiting body on the ellipse. The transformation from Keplerian elements to ECI coordinates possesses singularities at $e = 0$ and $i = 0, \pi$. Consequently, the propagation of nearly circular orbits or orbits of small inclination (and their uncertainties) lead to numerical instabilities. Instead, the *equinoctial orbital elements* [1] (a, h, k, p, q, ℓ) are a preferred set of orbital element coordinates

because they remove the zero-eccentricity and zero-inclination singularities. Although they do not possess a simple physical interpretation, their definition in terms of the Keplerian elements is straightforward:

$$a = a, \quad h = e\sin(\omega+\Omega), \quad k = e\cos(\omega+\Omega),$$
$$p = \tan\tfrac{i}{2}\sin\Omega, \quad q = \tan\tfrac{i}{2}\cos\Omega, \quad \ell = M+\omega+\Omega.$$

The transformation from equinoctial elements to ECI are provided in most astrodynamics textbooks (e.g., Montenbruck and Gill [19]). Models for the perturbing acceleration $\boldsymbol{a}_{pert}$ are also developed in this reference. The equinoctial elements are the default system of orbital element coordinates used in this manuscript.

The representation of the dynamical model (2.1) in coordinate systems other than ECI is straightforward to obtain using the chain rule. Indeed, if $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{r}, \dot{\boldsymbol{r}})$ denotes a coordinate transformation from ECI position-velocity coordinates $(\boldsymbol{r}, \dot{\boldsymbol{r}})$ to a coordinate system $\boldsymbol{u} \in \mathbb{R}^6$ (e.g., equinoctial orbital elements), then (2.1) is transformed to

$$\dot{\boldsymbol{u}} = \dot{\boldsymbol{u}}_{unpert} + \frac{\partial \boldsymbol{u}}{\partial \dot{\boldsymbol{r}}} \cdot \boldsymbol{a}_{pert}(\boldsymbol{r}, \dot{\boldsymbol{r}}, t), \tag{2.2}$$

where

$$\dot{\boldsymbol{u}}_{unpert} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{r}} \cdot \dot{\boldsymbol{r}} - \frac{\mu_\oplus}{r^3}\frac{\partial \boldsymbol{u}}{\partial \dot{\boldsymbol{r}}} \cdot \boldsymbol{r}. \tag{2.3}$$

If $\boldsymbol{u}$ is the vector of equinoctial orbital elements, then (2.3) simplifies to

$$\dot{\boldsymbol{u}}_{unpert} = \left(0, 0, 0, 0, 0, \sqrt{\mu_\oplus/a^3}\right)^T. \tag{2.4}$$

Therefore, the time evolution of a space object's equinoctial orbital element state (a, h, k, p, q, ℓ) under the assumption of unperturbed two-body dynamics is

$$a(t) = a_0, \quad h(t) = h_0, \quad k(t) = k_0, \quad p(t) = p_0, \quad q(t) = q_0, \quad \ell(t) = \ell_0 + n_0(t-t_0), \tag{2.5}$$

where $n_0 = \sqrt{\mu_\oplus/a_0^3}$ is the mean motion at the initial epoch.

We now argue that an equinoctial orbital element state can be regarded as a state defined on the six-dimensional cylinder $\mathbb{R}^5 \times \mathbb{S}$. The definition of the equinoctial orbital elements (a, h, k, p, q, ℓ) implicitly assumes closed (i.e., elliptical) orbits, which imposes the following constraints:

$$a > 0, \quad h^2 + k^2 < 1, \quad p, q \in \mathbb{R}, \quad \ell \in \mathbb{S}.$$

Though any real value can be assigned to the mean longitude coordinate ℓ, the angles ℓ and $\ell + 2\pi j$, for any integer $j \in \mathbb{Z}$, define the same location along the orbit; hence ℓ can be regarded as an angular coordinate defined on the circle $\mathbb{S}$. Mathematically, the first five elements (a, h, k, p, q) are not defined everywhere on $\mathbb{R}^5$. Physically, however, it can be argued that the sample space in these five coordinates can be approximated as all of $\mathbb{R}^5$. Indeed, the elements (a, h, k, p, q) describe the geometry of the (elliptical) orbit and the orientation of the orbit relative to the equatorial plane. In practice, the uncertainties in the orbital

geometry and orientation are sufficiently small so that the probability that an element is close to the constraint boundary is negligible. Moreover, under unperturbed dynamics, these five elements are conserved by Kepler's laws (see (2.5)), and, consequently, their respective uncertainties do not grow. That said, under the above assumptions, the manifold on which the elements (a, h, k, p, q) are defined can be approximated as all of $\mathbb{R}^5$, and the full six-dimensional equinoctial orbital element state space can be defined on the cylinder $\mathbb{R}^5 \times \mathbb{S}$. With the inclusion of perturbations in the dynamics (e.g., higher-order gravity terms, drag, and solar radiation), the first five elements only evolve with small periodic variations in time, and their uncertainties exhibit no long-term secular growth. Thus, the same cylindrical assumptions on the state space apply. These discussions also equally apply to other systems of orbital elements such as Poincaré orbital elements [14] and modified equinoctial orbital elements [27].

Though the uncertainties in the first five equinoctial elements exhibit only small periodic changes under two-body dynamics, it is the uncertainty along the semimajor axis coordinate a which causes the uncertainty along the mean longitude coordinate ℓ to grow without bound. In other words, as time progresses, one can generally maintain a good understanding of the geometry and orientation of the orbit, but confidence is gradually lost in the exact location of the object along its orbit. The growth in the uncertainty in ℓ can cause undesirable consequences if ℓ is incorrectly treated as an unbounded Cartesian variable or the uncertainty in ℓ is modeled as a Gaussian. With a Gaussian assumption imposed on the PDF in ℓ, one would assign different likelihoods to ℓ and $\ell + 2\pi j$ for different values of $j \in \mathbb{Z}$, even though the mean longitudes ℓ and $\ell + 2\pi j$ define the same location on the orbit. Additional insights are provided in the next subsection.

2.2. Examples of improper treatments of angular quantities. As described in the previous subsection, the mean longitude coordinate (i.e., the sixth equinoctial orbital element) is an angular coordinate in which the angles ℓ and $\ell + 2\pi j$ are identified (equivalent) for any integer $j \in \mathbb{Z}$. In other words, the mean longitude is a circular variable, and, as such, a rigorous treatment should not treat it as an unbounded real-valued Cartesian variable. In many cases (and with proper branch cuts defined), it is practical to treat the mean anomaly as real-valued, allowing for distributions to be represented as Gaussians, for example. The drawback of this approach is that the resulting statistics can depend on the choice of the integer "k."

For example, suppose one wishes to compute the conventional average $\bar{\theta}$ of two angles $\theta_1 = \pi/6$ and $\theta_2 = -\pi/6$. On one hand, $\bar{\theta} = \frac{1}{2}(\theta_1 + \theta_2) = 0$. On the other hand, since $\theta_2 = -\pi/6$ and $\theta_2 = 11\pi/6$ define the same location on the circle, using this equivalent value of θ_2 in the definition of the average would yield $\bar{\theta} = \pi$. Thus, in order for the conventional average to be well defined, the statistic

$$\bar{\theta} = \tfrac{1}{2}\big[(\theta_1 + 2\pi j_1) + (\theta_2 + 2\pi j_2)\big] = \tfrac{1}{2}(\theta_1 + \theta_2) + \pi(j_1 + j_2)$$

must be *independent* of the choice of $j_1, j_2 \in \mathbb{Z}$. Clearly this is not the case for the simple example studied above; different values of j_1 and j_2 can define averages at opposite sides of the circle. The unscented transform used in the UKF can also exhibit similar side effects when used in equinoctial orbital element space (especially if the mean longitude components of the sigma points are sufficiently dispersed) because it requires one to compute a weighted average

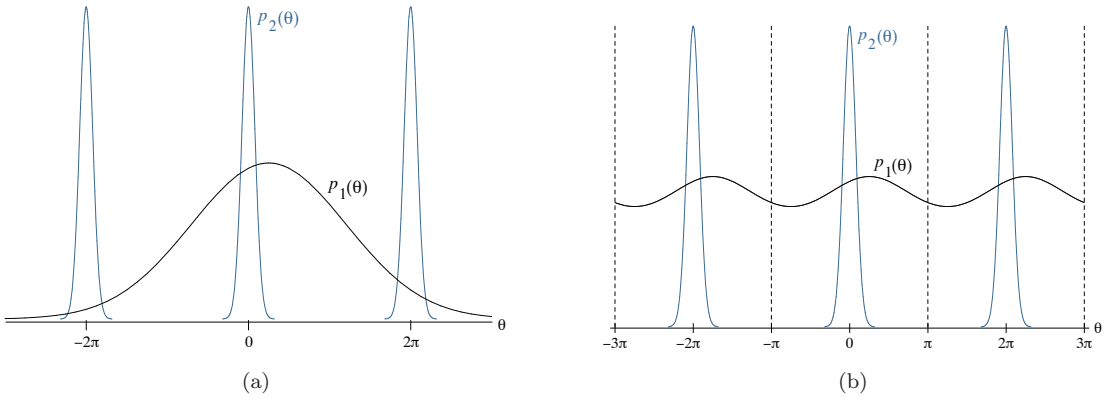


Figure 1. *Depiction of a prior PDF $p_1(\theta)$ and an updated PDF $p_2(\theta)$ in a circular variable θ when* (a) *the prior is misrepresented as a nonperiodic distribution and* (b) *the prior is represented correctly as a circular distribution. If the prior is fused with the information from the update using a standard Kalman filter, ambiguity would arise in* (a) *due to the need to select the correct 2π shift for the update. In* (b)*, this ambiguity is removed by representing both the prior and the update as circular distributions and using a fusion operation that is commensurate with the theory of directional statistics.*

of angular (mean longitude) components.

A more striking example of the misrepresentation of a circular variable as an unbounded Cartesian variable is demonstrated in Figure 1(a). In this simple example, consider a one-dimensional angular state space with two independent states: a "prior" and an "update" with respective PDFs $p_1(\theta)$ and $p_2(\theta)$. The prior PDF (depicted by the black curve) is diffuse (i.e., the uncertainty is large), and, furthermore, it is incorrectly modeled as an unbounded nonperiodic distribution. The updated PDF (depicted by one of the blue curves) is properly modeled as a circular distribution possessing a mean of zero and a very small (circular) variance (uncertainty). Any one of the blue curves can represent the PDF of the update since they are all equivalent up to an integer 2π shift. Within the Bayesian nonlinear filter correction step, one "fuses" the prior with the information from the update. In such an application, the correction step yields a "fused" PDF $p_f(\theta)$ given by

$$p_f(\theta) = \frac{1}{c}\, p_1(\theta)\, p_2(\theta), \quad c = \int_I p_1(\theta)\, p_2(\theta)\, \mathrm{d}\theta, \tag{2.6}$$

where $I = (-\infty, \infty)$. In a filter correction step that requires a Gaussian representation of both the prior and the update (e.g., in the traditional Kalman filter), one would need to "convert" the circular distribution $p_2(\theta)$ to a Gaussian. Ambiguity would arise in this procedure due to the need to choose the "correct" 2π shift; i.e., the selection of an integer k such that the mean of $p_2(\theta)$ is $2\pi k$. The fused PDF and the constant c in (2.6) would depend on this choice of k. Indeed, a miscomputed normalization constant c can have severe consequences within a tracking system, as analogous expressions for c appear in the likelihood ratios for scoring the association of one report to another [22]. Figure 1(b) depicts how these ambiguity issues can be resolved by turning to the theory of directional statistics and properly modeling both the prior and update as circular distributions. Further analysis is provided in the next subsection.

2.3. Von Mises PDF. The source of the ambiguity in computing the fused PDF (2.6) in the example of the previous subsection lies in the incorrect treatment of an angular variable as an unbounded Cartesian real-valued variable. These problems can be rectified by representing the state PDF not necessarily on a Cartesian space $\mathbb{R}^n$ but instead on the manifold that more accurately describes the global topology of the underlying state space. The most important change needed for (equinoctial) orbital element states is the representation of the uncertainty in the angular coordinate (mean longitude) as a PDF on the circle $\mathbb{S}$ so that the joint PDF in the six orbital elements defines a distribution on the cylinder $\mathbb{R}^5 \times \mathbb{S}$. In order to do so, PDFs defined on the circle are required.

A function $p : \mathbb{R} \to \mathbb{R}$ is a probability density function (PDF) on the circle $\mathbb{S}$ if and only if [17, sect. 3.2]

1. $p(\theta) \geq 0$ almost everywhere on $(-\infty, \infty)$,
2. $p(\theta + 2\pi) = p(\theta)$ almost everywhere on $(-\infty, \infty)$,
3. $\int_{-\pi}^{\pi} p(\theta)\, \mathrm{d}\theta = 1$.

One example satisfying the above conditions is the *von Mises distribution*, which provides an analogy of a Gaussian distribution defined on a circle. The von Mises PDF in the angular variable θ is defined by [16, 17]

$$\mathcal{VM}(\theta; \alpha, \kappa) = \frac{\mathrm{e}^{\kappa \cos(\theta - \alpha)}}{2\pi I_0(\kappa)}, \tag{2.7}$$

where I_0 is the modified Bessel function of the first kind of order 0. The parameters α and κ are measures of location and concentration. As $\kappa \to 0^+$, the von Mises distribution tends to a uniform distribution. For large κ, the von Mises distribution becomes concentrated about the angle α and approaches a Gaussian distribution in θ with mean α and variance $1/\kappa$. An algebraically equivalent expression of (2.7) which is numerically stable for large values of κ is

$$\mathcal{VM}(\theta; \alpha, \kappa) = \frac{\mathrm{e}^{-2\kappa \sin^2 \frac{1}{2}(\theta - \alpha)}}{2\pi \mathrm{e}^{-\kappa} I_0(\kappa)}. \tag{2.8}$$

The von Mises distribution (2.8) is used in the next section to construct the GVM distribution defined on the cylinder $\mathbb{R}^n \times \mathbb{S}$.

This section concludes by discussing how the von Mises distribution can resolve the problems in the previous subsection concerning the averaging and fusion of angular quantities. Suppose $\theta_1, \ldots, \theta_N$ is a sample of independent observations coming from a von Mises distribution. Then, the maximum likelihood estimate of the location parameter α is

$$\hat{\alpha} = \arg\left(\frac{1}{N} \sum_{j=1}^{N} \mathrm{e}^{\mathrm{i}\theta_j} \right). \tag{2.9}$$

Unlike the conventional average $(\theta_1 + \cdots + \theta_N)/N$, the "average" $\hat{\alpha}$ is *independent* of any 2π shift in the angles θ_j. In the fusion example, if the prior and update PDFs are von Mises distributions, i.e.,

$$p_1(\theta) = \mathcal{VM}(\theta; \alpha_1, \kappa_1), \quad p_2(\theta) = \mathcal{VM}(\theta; \alpha_2, \kappa_2),$$

then the fused PDF (2.6) can be formed unambiguously, and $p_f(\theta) = p_f(\theta+2\pi j)$ for any $j \in \mathbb{Z}$. As seen in Figure 1(b), there is no longer any ambiguity in how to choose the "correct" update PDF $p_2(\theta)$, since both $p_1(\theta)$ and $p_2(\theta)$ are properly defined circular PDFs. The integration interval I in the computation of the normalization constant c in (2.6) is any interval of length 2π; i.e., $I = (a, a + 2\pi)$ for some $a \in \mathbb{R}$. In other words, the value of c is independent of the choice of a.

3. Construction of the GVM distribution. This section defines the Gauss von Mises (GVM) distribution used to characterize the uncertainty in a space object's orbital state. Later, in subsection 3.1, the connection between the GVM distribution and a related family of distributions proposed by Mardia and Sutton [18] is discussed.

The proposed GVM distribution is not defined as a function of other random variables with specified PDFs. Instead, its construction is based on satisfying the following requirements.

1. The GVM family of multivariate PDFs is defined on the $n+1$-dimensional cylindrical manifold $\mathbb{R}^n \times \mathbb{S}$.
2. The GVM distribution has a sufficiently general parameter set which can model nonzero higher-order cumulants beyond a usual "mean" and "covariance."
3. The GVM distribution can account for correlation between the Cartesian random vector $\boldsymbol{x} \in \mathbb{R}^n$ and the circular random variable $\theta \in \mathbb{S}$.
4. The level sets of the GVM distribution are generally "banana" or "boomerang" shaped (as motivated in the introduction) so that a more accurate characterization of the uncertainty in a space object's state in equinoctial orbital elements can be achieved.
5. The GVM distribution reduces to a multivariate Gaussian PDF in a suitable limit.
6. The GVM distribution permits a tractable implementation of the general Bayesian nonlinear filter and other applications needed to support advanced SSA.

To clarify the first requirement, a function $p : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a PDF on the cylinder $\mathbb{R}^n \times \mathbb{S}$ if and only if

1. $p(\boldsymbol{x}, \theta) \geq 0$ almost everywhere on $\mathbb{R}^n \times \mathbb{R}$,
2. $p(\boldsymbol{x}, \theta + 2\pi) = p(\boldsymbol{x}, \theta)$ almost everywhere on $\mathbb{R}^n \times \mathbb{R}$,
3. $\int_{\mathbb{R}^n} \int_{-\pi}^{\pi} p(\boldsymbol{x}, \theta)\, \mathrm{d}\theta\, \mathrm{d}\boldsymbol{x} = 1$.

This definition extends the definition of a PDF defined on a circle (see page 283 or section 3.2 of [17]). One example satisfying the above conditions is

$$p(\boldsymbol{x}, \theta) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P})\, \mathcal{VM}(\theta; \alpha, \kappa),$$

where $\mathcal{VM}(\theta; \alpha, \kappa)$ is the von Mises PDF defined by (2.8) and $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P})$ is the multivariate Gaussian PDF given by

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P}) = \frac{1}{\sqrt{\det(2\pi\mathbf{P})}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \mathbf{P}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right], \tag{3.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\mathbf{P}$ is an $n \times n$ symmetric positive-definite (covariance) matrix. This simple example, though used as a starting point in constructing the GVM distribution, does not satisfy all of the requirements listed earlier. In particular, the desired family of multivariate PDFs needs to model correlation between $\boldsymbol{x}$ and θ and have level sets possessing a distinctive banana or boomerang shape.

To motivate the construction of the GVM distribution, consider two random vectors $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$ whose joint PDF is Gaussian:

$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{P}_{yx}^T \\ \mathbf{P}_{yx} & \mathbf{P}_{yy} \end{bmatrix} \right). \tag{3.2}$$

Using the definition of conditional probability and the *Schur complement decomposition* [8], the joint PDF (3.2) can be expressed as

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{y}) &= p(\boldsymbol{x})\, p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_x, \mathbf{P}_{xx})\, \mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}_y + \mathbf{P}_{yx}\mathbf{P}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \mathbf{P}_{yy} - \mathbf{P}_{yx}\mathbf{P}_{xx}^{-1}\mathbf{P}_{yx}^T) \\ &\equiv \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_x, \mathbf{P}_{xx})\, \mathcal{N}(\boldsymbol{y}; \boldsymbol{g}(\boldsymbol{x}), \mathbf{Q}), \end{aligned} \tag{3.3}$$

where

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{\mu}_y + \mathbf{P}_{yx}\mathbf{P}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \quad \mathbf{Q} = \mathbf{P}_{yy} - \mathbf{P}_{yx}\mathbf{P}_{xx}^{-1}\mathbf{P}_{yx}^T.$$

One very powerful observation resulting from this Schur complement decomposition is that (3.3) defines a PDF in $(\boldsymbol{x}, \boldsymbol{y})$ for *any* (analytic) function $\boldsymbol{g}(\boldsymbol{x})$ and *any* symmetric positive-definite matrix $\mathbf{Q}$. In particular, if $\boldsymbol{y}$ is univariate and is relabeled as θ, then

$$p(\boldsymbol{x}, \theta) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_x, \mathbf{P}_{xx})\, \mathcal{N}(\theta; \Theta(\boldsymbol{x}), \kappa)$$

defines a PDF on $\mathbb{R}^n \times \mathbb{R}$ for any analytic function $\Theta : \mathbb{R}^n \to \mathbb{R}$ and any positive scalar κ. To make it robust for a *circular variable* $\theta \in \mathbb{S}$ and hence define a PDF on the cylinder $\mathbb{R}^n \times \mathbb{S}$, we replace the Gaussian PDF in θ by the von Mises PDF in θ:

$$p(\boldsymbol{x}, \theta) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_x, \mathbf{P}_{xx})\, \mathcal{VM}(\theta; \Theta(\boldsymbol{x}), \kappa). \tag{3.4}$$

The definition of the GVM distribution fixes the specific form of the function $\Theta(\boldsymbol{x})$ in (3.4) so that the PDF can model nonzero higher-order cumulants (i.e., the banana shape of the level sets) but is not overly complicated so as the make the resulting Bayesian filter prediction and correction steps intractable.

Definition 3.1 (Gauss von Mises (GVM) distribution). *The random variables $(\boldsymbol{x}, \theta) \in \mathbb{R}^n \times \mathbb{S}$ are said to be jointly distributed as a Gauss von Mises (GVM) distribution if and only if their joint PDF has the form*

$$p(\boldsymbol{x}, \theta) = \mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) \equiv \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P})\, \mathcal{VM}(\theta; \Theta(\boldsymbol{x}), \kappa),$$

where

$$\begin{aligned} \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P}) &= \frac{1}{\sqrt{\det(2\pi\mathbf{P})}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \mathbf{P}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right], \\ \mathcal{VM}(\theta; \Theta(\boldsymbol{x}), \kappa) &= \frac{1}{2\pi \mathrm{e}^{-\kappa} I_0(\kappa)} \exp\left[-2\kappa \sin^2 \frac{1}{2}\big(\theta - \Theta(\boldsymbol{x})\big) \right], \end{aligned}$$

and

$$\Theta(\boldsymbol{x}) = \alpha + \boldsymbol{\beta}^T \boldsymbol{z} + \tfrac{1}{2}\boldsymbol{z}^T \boldsymbol{\Gamma} \boldsymbol{z}, \quad \boldsymbol{z} = \mathbf{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \quad \mathbf{P} = \mathbf{A}\mathbf{A}^T.$$

The parameter set $(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ *is subject to the following constraints:* $\boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{P}$ *is an* $n \times n$ *symmetric positive-definite matrix,* $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\Gamma}$ *is an* $n \times n$ *symmetric matrix, and* $\kappa \geq 0$. *The matrix* $\mathbf{A}$ *in the definition of the normalized variable* $\boldsymbol{z}$ *is the lower-triangular Cholesky factor* [8] *of the parameter matrix* $\mathbf{P}$. *The notation* $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ *denotes that* $(\boldsymbol{x}, \theta)$ *are jointly distributed as a GVM distribution with the specified parameter set.*

It is noted that the function $\Theta(\boldsymbol{x})$ appearing in the GVM distribution is an inhomogeneous quadratic in $\boldsymbol{x}$ or, equivalently, the normalized variable $\boldsymbol{z}$. We use the normalized variable $\boldsymbol{z}$ in the definition to simplify the ensuing mathematics and so that the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ are dimensionless. Additionally, the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ model correlation between $\boldsymbol{x}$ and θ. The parameter matrix $\boldsymbol{\Gamma}$ can be tuned to give the level sets of the GVM distribution their distinctive banana or boomerang shape.

3.1. Relationship to Mardia and Sutton. A related distribution defined on a cylindrical manifold proposed by Mardia and Sutton [18] can be recovered starting from (3.3). If $\boldsymbol{x}$ is univariate and is relabeled as θ, then

$$p(\boldsymbol{y}, \theta) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{g}(\theta), \mathbf{Q})\, \mathcal{N}(\theta; \mu_\theta, P_{\theta\theta})$$

defines a PDF on $\mathbb{R}^m \times \mathbb{R}$ for any analytic function $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^m$. A PDF defined on the cylinder $\mathbb{R}^m \times \mathbb{S}$ can be obtained by replacing the Gaussian PDF in θ by the von Mises PDF in θ (and relabeling the other parameters) to yield

$$p(\boldsymbol{y}, \theta) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{g}(\theta), \mathbf{P})\, \mathcal{VM}(\theta; \alpha, \kappa). \tag{3.5}$$

One specific choice for the function $\boldsymbol{g}(\theta)$ is

$$\boldsymbol{g}(\theta) = \boldsymbol{a}_0 + \boldsymbol{a}_1 \cos\theta + \boldsymbol{b}_1 \sin\theta \tag{3.6}$$

for some parameter vectors $\boldsymbol{a}_0, \boldsymbol{a}_1, \boldsymbol{b}_1 \in \mathbb{R}^m$.

The PDF (3.5) with $\boldsymbol{g}(\theta)$ specified by (3.6) is a generalization of the bivariate distribution defined on the two-dimensional cylinder $\mathbb{R} \times \mathbb{S}$ proposed by Mardia and Sutton [18]. Their derivation also differs from that presented here in the sense that the authors start with a trivariate Gaussian distribution and then "project it" onto the cylinder.

Though the Mardia–Sutton distribution can be used in place of the GVM distribution in estimation and other SSA functions, it is noted that the parameter set in the former provides less control over the magnitude of the higher-order cumulants and the bending of the banana- or boomerang-shaped level sets. For these reasons, the GVM distribution is preferred since it overcomes these limitations.

4. Elementary properties. This section lists the important mathematical and statistical properties of the GVM distribution based on the definition stated in section 3. These properties are used in subsequent sections of the paper in the development of the uncertainty propagation algorithm using the GVM distribution. Unless stated otherwise, the derivation of the results follow from elementary means.

4.1. Mode. The GVM distribution is *unimodal* on $\mathbb{R}^n \times \mathbb{S}$ with

$$(\boldsymbol{\mu}, \alpha) = \arg\max_{\boldsymbol{x},\theta} \mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa).$$

4.2. Characteristic function. The characteristic function of the GVM distribution, i.e.,

$$\begin{aligned}\varphi_{\mathcal{GVM}}(\boldsymbol{\xi}, m; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) &\equiv \mathrm{E}\big[\mathrm{e}^{\mathrm{i}(\boldsymbol{\xi}^T\boldsymbol{x} + m\theta)}\big] \\ &= \int_{\mathbb{R}^n}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(\boldsymbol{\xi}^T\boldsymbol{x} + m\theta)}\, \mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)\, \mathrm{d}\theta\, \mathrm{d}\boldsymbol{x}\end{aligned}$$

for $\boldsymbol{\xi} \in \mathbb{R}^n$ and $m \in \mathbb{Z}$, is

$$\begin{aligned}\varphi_{\mathcal{GVM}}(\boldsymbol{\xi}, m; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) &= \frac{1}{\sqrt{\det(\mathbf{I} - \mathrm{i}m\boldsymbol{\Gamma})}} \frac{I_{|m|}(\kappa)}{I_0(\kappa)} \\ &\times \exp\left[\mathrm{i}(\boldsymbol{\mu}^T\boldsymbol{\xi} + m\alpha) - \frac{1}{2}(\mathbf{A}^T\boldsymbol{\xi} + m\boldsymbol{\beta})^T(\mathbf{I} - \mathrm{i}m\boldsymbol{\Gamma})^{-1}(\mathbf{A}^T\boldsymbol{\xi} + m\boldsymbol{\beta})\right],\end{aligned} \tag{4.1}$$

where $\mathbf{I}$ denotes the $n \times n$ identity matrix and I_p is the modified Bessel function of the first kind of order p.

4.3. Low-order moments. If $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$, then

$$\begin{gathered}\mathrm{E}[\boldsymbol{x}] = \boldsymbol{\mu}, \\ \mathrm{E}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T] = \mathbf{P}, \\ \mathrm{E}[\mathrm{e}^{\mathrm{i}\theta}] = \frac{1}{\sqrt{\det(\mathbf{I} - \mathrm{i}\boldsymbol{\Gamma})}} \frac{I_1(\kappa)}{I_0(\kappa)} \exp\left[\mathrm{i}\alpha - \frac{1}{2}\boldsymbol{\beta}^T(\mathbf{I} - \mathrm{i}\boldsymbol{\Gamma})^{-1}\boldsymbol{\beta}\right], \\ \mathrm{E}[\mathrm{e}^{\mathrm{i}\theta}\,\boldsymbol{z}] = \mathrm{i}(\mathbf{I} - \mathrm{i}\boldsymbol{\Gamma})^{-1}\boldsymbol{\beta}\,\mathrm{E}[\mathrm{e}^{\mathrm{i}\theta}], \\ \mathrm{E}[\mathrm{e}^{\mathrm{i}\theta}\,\boldsymbol{z}\boldsymbol{z}^T] = \big[(\mathbf{I} - \mathrm{i}\boldsymbol{\Gamma})^{-1} - (\mathbf{I} - \mathrm{i}\boldsymbol{\Gamma})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^T(\mathbf{I} - \mathrm{i}\boldsymbol{\Gamma})^{-1}\big]\,\mathrm{E}[\mathrm{e}^{\mathrm{i}\theta}],\end{gathered}$$

where $\boldsymbol{z} = \mathbf{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ and $\mathbf{P} = \mathbf{A}\mathbf{A}^T$. These results follow from (4.1).

4.4. Invariance property. If $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ defined on $\mathbb{R}^n \times \mathbb{S}$ and

$$\tilde{\boldsymbol{x}} = \mathbf{L}\boldsymbol{x} + \boldsymbol{d}, \quad \tilde{\theta} = \theta + a + \boldsymbol{b}^T\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}^T\mathbf{C}\boldsymbol{x}, \tag{4.2}$$

where $\mathbf{L}$ is an $m \times n$ matrix (with $m \le n$), $\boldsymbol{d} \in \mathbb{R}^m$, $a \in \mathbb{R}$, $\boldsymbol{b} \in \mathbb{R}^n$, and $\mathbf{C}$ is a symmetric $n \times n$ matrix, then

$$(\tilde{\boldsymbol{x}}, \tilde{\theta}) \sim GVM(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{P}}, \tilde{\alpha}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Gamma}}, \tilde{\kappa})$$

defined on $\mathbb{R}^m \times \mathbb{S}$, where

$$\begin{gathered}\tilde{\boldsymbol{\mu}} = \mathbf{L}\boldsymbol{\mu} + \boldsymbol{d}, \quad \tilde{\mathbf{P}} = \mathbf{L}\mathbf{P}\mathbf{L}^T, \quad \tilde{\alpha} = \alpha + a + \boldsymbol{b}^T\boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{\mu}^T\mathbf{C}\boldsymbol{\mu}, \\ \tilde{\boldsymbol{\beta}} = \mathbf{Q}^T[\boldsymbol{\beta} + \mathbf{A}^T(\mathbf{C}\boldsymbol{\mu} + \boldsymbol{b})], \quad \tilde{\boldsymbol{\Gamma}} = \mathbf{Q}^T(\boldsymbol{\Gamma} + \mathbf{A}^T\mathbf{C}\mathbf{A})\mathbf{Q}, \quad \tilde{\kappa} = \kappa.\end{gathered} \tag{4.3}$$

In these equations, $\mathbf{Q}$ is an $n \times m$ matrix with mutually orthonormal columns, and $\mathbf{R}$ is an $m \times m$ positive-definite upper-triangular matrix such that $\mathbf{Q}\mathbf{R} = (\mathbf{L}\mathbf{A})^T$. (The $\mathbf{Q}$ and $\mathbf{R}$

matrices can be computed by performing a "thin" QR factorization [8, sect. 5.2] of $(\mathbf{LA})^T$.) Additionally, the lower-triangular Cholesky factor $\tilde{\mathbf{A}}$ such that $\tilde{\mathbf{P}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$ is $\tilde{\mathbf{A}} = \mathbf{R}^T$.

This result, which states that a GVM distribution remains a GVM distribution under transformations of the form (4.2), follows upon computing the characteristic function of the transformed random variables $(\tilde{\boldsymbol{x}}, \tilde{\theta})$ and then identifying it with the characteristic function (4.1) possessing the parameters defined in (4.3).

4.5. Transformation to canonical form. The transformation

$$\boldsymbol{z} = \mathbf{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \quad \phi = \theta - \alpha - \boldsymbol{\beta}^T\boldsymbol{z} - \tfrac{1}{2}\boldsymbol{z}^T\boldsymbol{\Gamma}\boldsymbol{z} \tag{4.4}$$

reduces $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ to the *canonical or standardized GVM distribution* with PDF

$$p(\boldsymbol{z}, \phi) = \mathcal{GVM}(\boldsymbol{z}, \phi; \mathbf{0}, \mathbf{I}, 0, \mathbf{0}, \mathbf{0}, \kappa) = \mathcal{N}(\boldsymbol{z}; \mathbf{0}, \mathbf{I})\, \mathcal{VM}(\phi; 0, \kappa).$$

This result follows by noting that the transformation (4.4) is a special case of (4.2).

4.6. Conversion between GVM and Gaussian distributions. As $\kappa \to \infty$ and $\|\boldsymbol{\Gamma}\| \to 0$, a GVM distribution in $(\boldsymbol{x}, \theta)$ becomes a Gaussian distribution in $\mathbf{x} = (\boldsymbol{x}, \theta)$. Under these conditions,

$$\mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) \approx \mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}), \tag{4.5}$$

where

$$\mathbf{m} = \begin{bmatrix} \boldsymbol{\mu} \\ \alpha \end{bmatrix}, \quad \mathbf{P} = \mathbf{A}\mathbf{A}^T, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \boldsymbol{\beta}^T & 1/\sqrt{\kappa} \end{bmatrix}. \tag{4.6}$$

The approximation (4.5) is justified by first noting that, for fixed $\boldsymbol{x}$, the von Mises distribution $\mathcal{VM}(\theta; \Theta(\boldsymbol{x}), \kappa)$ becomes concentrated at the point $\theta = \Theta(\boldsymbol{x})$ as $\kappa \to \infty$. Thus, for large κ,

$$\begin{aligned} \mathcal{VM}(\theta; \Theta(\boldsymbol{x}), \kappa) &= \frac{1}{2\pi \mathrm{e}^{-\kappa} I_0(\kappa)} \exp\left[-2\kappa \sin^2 \frac{1}{2}\big(\theta - \Theta(\boldsymbol{x})\big)\right] \\ &\approx \sqrt{\frac{\kappa}{2\pi}} \exp\left[-\frac{1}{2}\kappa\big(\theta - \Theta(\boldsymbol{x})\big)^2\right] = \mathcal{N}(\theta; \Theta(\boldsymbol{x}), 1/\kappa), \end{aligned}$$

noting that $I_0(\kappa) \sim \mathrm{e}^{\kappa}/\sqrt{2\pi\kappa}$ as $\kappa \to \infty$ and $\sin^2 \frac{1}{2}\phi = \frac{1}{4}\phi^2$ as $\phi \to 0$; hence

$$\mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) \approx \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{P})\, \mathcal{N}(\theta; \Theta(\boldsymbol{x}), 1/\kappa). \tag{4.7}$$

As $\|\boldsymbol{\Gamma}\| \to 0$, $\Theta(\boldsymbol{x})$ tends to a linear function in $\boldsymbol{x}$, thereby reducing the right-hand side of (4.7) to a Gaussian in $\boldsymbol{x}$ and θ.

The expressions for the mean $\mathbf{m}$ and covariance $\mathbf{P}$ in (4.6) of the approximating Gaussian in $\mathbf{x} = (\boldsymbol{x}, \theta)$ are derived as follows. First, define

$$q_{\mathcal{GVM}}(\boldsymbol{x}, \theta) = -\ln \mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa), \quad q_{\mathcal{N}}(\mathbf{x}) = -\ln \mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}).$$

The approximating Gaussian in (4.5) is the "osculating Gaussian" defined as the Gaussian which is tangent to the GVM distribution at the mode. Specifically, the mean $\mathbf{m}$ is the mode of the GVM distribution, as specified in (4.6), and the covariance $\mathbf{P}$ is obtained by demanding equality between the second-order partial derivatives of $q_{\mathcal{GVM}}$ and $q_{\mathcal{N}}$ evaluated at the mode. Indeed,

$$\left.\frac{\partial^2 q_{\mathcal{N}}}{\partial \mathbf{x}\, \partial \mathbf{x}^T}\right|_{\mathbf{x}=\mathbf{m}} = \mathbf{P}^{-1}, \qquad \left.\frac{\partial^2 q_{\mathcal{GVM}}}{\partial \boldsymbol{x}\, \partial \boldsymbol{x}^T}\right|_{(\boldsymbol{x},\theta)=(\boldsymbol{\mu},\alpha)} = \mathbf{A}^{-T}(\mathbf{I} + \kappa \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{A}^{-1},$$
$$\left.\frac{\partial^2 q_{\mathcal{GVM}}}{\partial \theta\, \partial \boldsymbol{x}}\right|_{(\boldsymbol{x},\theta)=(\boldsymbol{\mu},\alpha)} = -\kappa \mathbf{A}^{-T}\boldsymbol{\beta}, \qquad \left.\frac{\partial^2 q_{\mathcal{GVM}}}{\partial \theta^2}\right|_{(\boldsymbol{x},\theta)=(\boldsymbol{\mu},\alpha)} = \kappa;$$

hence

$$\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{A}^{-T}(\mathbf{I} + \kappa \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{A}^{-1} & -\kappa \mathbf{A}^{-T}\boldsymbol{\beta} \\ -\kappa \boldsymbol{\beta}^T \mathbf{A}^{-1} & \kappa \end{bmatrix}. \tag{4.8}$$

Inverting (4.8) using the block matrix inversion lemma yields

$$\mathbf{P} = \begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{A}\boldsymbol{\beta} \\ \boldsymbol{\beta}^T \mathbf{A}^T & \boldsymbol{\beta}^T\boldsymbol{\beta} + 1/\kappa \end{bmatrix}. \tag{4.9}$$

Finally, computing the lower-triangular Cholesky factor $\mathbf{A}$ of the covariance (4.9) results in the expression in (4.6).

In summary, (4.5) and (4.6) dictate how one can convert to and from a GVM distribution and a Gaussian. The approximation of a Gaussian distribution in $\mathbf{x} = (\boldsymbol{x}, \theta)$ by a GVM distribution with $\boldsymbol{\Gamma} = \mathbf{0}$ becomes exact in the limit as the standard deviation in θ tends to zero. The approximation of a GVM distribution by a Gaussian becomes exact as $\kappa \to \infty$ and $\|\boldsymbol{\Gamma}\| \to 0$. In any case, the equations provide an approximation of a GVM distribution by the osculating or tangent Gaussian distribution at the modal point.

4.7. Mahalanobis von Mises statistic. Given $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$, the *Mahalanobis von Mises statistic* is defined by

$$M(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) \equiv (\boldsymbol{x} - \boldsymbol{\mu})^T \mathbf{P}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + 4\kappa \sin^2 \tfrac{1}{2}\big(\theta - \Theta(\boldsymbol{x})\big), \tag{4.10}$$

where $\Theta(\boldsymbol{x}) = \alpha + \boldsymbol{\beta}^T \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^T \boldsymbol{\Gamma} \boldsymbol{z}$, $\boldsymbol{z} = \mathbf{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$, and $\mathbf{P} = \mathbf{A}\mathbf{A}^T$. For large κ, $M \mathrel{\dot{\sim}} \chi^2(n+1)$, where $\chi^2(\nu)$ is the chi-square distribution with ν degrees of freedom and "$\mathrel{\dot{\sim}}$" means "approximately distributed." The statistic (4.10) is analogous to the Mahalanobis distance [15] for a multivariate Gaussian random vector $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \mathbf{P})$. Indeed, the first term in (4.10) is precisely this Mahalanobis distance. Like the classical Mahalanobis distance, the Mahalanobis von Mises statistic provides a mechanism for testing if some realization $(\boldsymbol{x}_*, \theta_*)$ from a GVM distribution is statistically significant.

The derivation of the distribution of M given that $(\boldsymbol{x}, \theta)$ are jointly distributed as a GVM distribution is as follows. Applying the transformation (4.4) to canonical form yields

$$M = \boldsymbol{z}^T \boldsymbol{z} + 4\kappa \sin^2 \tfrac{1}{2}\phi,$$

where now $\boldsymbol{z} \sim N(\mathbf{0}, \mathbf{I})$ and $\phi \sim VM(0, \kappa)$ with $\boldsymbol{z}$ and ϕ independent. Therefore, $M = \chi^2(n) + Y_\kappa$, where $Y_\kappa \equiv 4\kappa \sin^2 \frac{1}{2}\phi$ and is independent of the $\chi^2(n)$ random variable. It follows from elementary probability theory that the PDF of Y_κ is

$$p_{Y_\kappa}(y) = \begin{cases} \dfrac{\mathrm{e}^{-y/2}}{\pi \mathrm{e}^{-\kappa} I_0(\kappa) \sqrt{y(4\kappa - y)}}, & 0 < y < 4\kappa, \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

Noting that $I_0(\kappa) \sim \mathrm{e}^\kappa / \sqrt{2\pi\kappa}$ and $\sqrt{y(4\kappa - y)} \sim 2\sqrt{\kappa y}$ as $\kappa \to \infty$, it follows that

$$p_{Y_\kappa}(y) \sim \frac{\mathrm{e}^{-y/2}}{\sqrt{2\pi y}} = p_{\chi^2(1)}(y).$$

In other words, the PDF of Y_κ converges (pointwise) to the PDF of a chi-square random variable with one degree of freedom as $\kappa \to \infty$. It can also be shown that the cumulative distribution function (CDF) of Y_κ converges uniformly to the CDF of $\chi^2(1)$ as $\kappa \to \infty$. Therefore, for large κ, $Y_\kappa \dot{\sim} \chi^2(1)$ and hence $M \dot{\sim} \chi^2(n+1)$.

In principle, one can derive the exact PDF of M using the standard change of variables theorem in conjunction with the PDFs of the chi-square distribution and (4.11), though an analytic expression for the resulting convolution is intractable. In practice and in many of the applications considered in SSA, κ is sufficiently large so that the approximation of Y_κ by a $\chi^2(1)$ random variable (and hence M by $\chi^2(n+1)$) is justified.

5. GVM quadrature.

The Julier–Uhlmann unscented transform [13] or the more general framework of Gauss–Hermite quadrature [6] enables one to compute the expected value of a nonlinear transformation of a multivariate Gaussian random vector. In this section, these methodologies are extended to enable the computation of the expected value of a nonlinear transformation of a GVM random vector, thereby providing a general framework for *GVM quadrature*. The quadrature formulas are subsequently used in the prediction step of the GVM filter developed in the next section. The GVM quadrature framework is also useful for extracting supplemental statistics from a GVM distribution. For example, if a space object's orbital state is represented as a GVM distribution in equinoctial orbital element space, one can compute the expected value of the object's Cartesian position and velocity or the covariance in its position and velocity.

Given $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ and a function $f : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}$, we seek an approximation to

$$\mathrm{E}[f(\boldsymbol{x}, \theta)] = \int_{\mathbb{R}^n} \int_{-\pi}^{\pi} \mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)\, f(\boldsymbol{x}, \theta)\, \mathrm{d}\theta\, \mathrm{d}\boldsymbol{x}. \tag{5.1}$$

As in classical Gaussian quadrature, the framework of GVM quadrature approximates (5.1) as a weighted sum of function values at specified points:

$$\int_{\mathbb{R}^n} \int_{-\pi}^{\pi} \mathcal{GVM}(\boldsymbol{x}, \theta; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)\, f(\boldsymbol{x}, \theta)\, \mathrm{d}\theta\, \mathrm{d}\boldsymbol{x} \approx \sum_{i=1}^{N} w_{\sigma_i} f(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i}). \tag{5.2}$$

The set of *quadrature nodes*, sometimes called sigma points, $\{(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i})\}_{i=1}^N$ and corresponding *quadrature weights* $\{w_{\sigma_i}\}_{i=1}^N$ are chosen so that (5.2) is exact for a certain class of functions. In the derivation of the GVM quadrature weights and nodes, the Smolyak sparse grid paradigm [7, 25] is used so that, for a specified order of accuracy, the number of nodes increases only *polynomially* with the dimension n, thereby avoiding the so-called *curse of dimensionality*. In particular, the number of quadrature nodes (and hence function evaluations) increases *linearly* in the dimension n for the third-order rule derived in subsection 5.1. Though analogous higher-order sparse grid quadrature methods can be derived, such details are omitted in this paper. Note that in the six-dimensional setting of orbital state uncertainty propagation problems in SSA, the dimension is sufficiently high to warrant the use of such sparse grid methods. Indeed, a tensor product of a three-point quadrature rule in one dimension would yield $3^6 = 729$ quadrature nodes in six dimensions; this would be far too extensive to use in any real-time operational system.

5.1. Third-order method. Without loss of generality, the method is restricted to quadrature using the canonical GVM distribution as the weighting function:

$$\int_{\mathbb{R}^n} \int_{-\pi}^{\pi} \mathcal{N}(\boldsymbol{z}; \mathbf{0}, \mathbf{I})\, \mathcal{VM}(\phi; 0, \kappa)\, f(\boldsymbol{z}, \phi)\, \mathrm{d}\phi\, \mathrm{d}\boldsymbol{z} \approx \sum_{i=1}^{N} w_{\sigma_i} f(\boldsymbol{z}_{\sigma_i}, \phi_{\sigma_i}). \tag{5.3}$$

Indeed, if considering the general quadrature problem in (5.2), the nodes $(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i})$ can be recovered from the nodes $(\boldsymbol{z}_{\sigma_i}, \phi_{\sigma_i})$ of the canonical problem (5.3) through the transformation (4.4):

$$\boldsymbol{x}_{\sigma_i} = \boldsymbol{\mu} + \mathbf{A}\boldsymbol{z}_{\sigma_i}, \quad \theta_{\sigma_i} = \phi_{\sigma_i} + \alpha + \boldsymbol{\beta}^T \boldsymbol{z}_{\sigma_i} + \tfrac{1}{2} \boldsymbol{z}_{\sigma_i}^T \boldsymbol{\Gamma} \boldsymbol{z}_{\sigma_i}.$$

In what follows, the following notation is required. Let $\xi \in \mathbb{R}$, $\eta \in (-\pi, \pi]$, and

$$\begin{aligned}
\mathcal{N}^{00} &= \{(\boldsymbol{z}, \phi) \in \mathbb{R}^n \times \mathbb{S} \mid \boldsymbol{z} = \mathbf{0},\ \phi = 0\}, \\
\mathcal{N}^{\eta 0} &= \{(\boldsymbol{z}, \phi) \in \mathbb{R}^n \times \mathbb{S} \mid \boldsymbol{z} = \mathbf{0},\ |\phi| = \eta\}, \\
\mathcal{N}_i^{\xi 0} &= \{(\boldsymbol{z}, \phi) \in \mathbb{R}^n \times \mathbb{S} \mid z_1 = \cdots = z_{i-1} = z_{i+1} = \cdots = z_n = 0,\ |z_i| = \xi,\ \phi = 0\}.
\end{aligned}$$

For the third-order method, it is demanded that the approximation (5.3) be exact for all functions of the form

$$f(\boldsymbol{z}, \phi) = \left(a_0 + \tfrac{1}{2} \boldsymbol{z}^T \mathbf{B}_0 \boldsymbol{z}\right) + a_1 \cos\phi + a_2 \cos 2\phi + g_o(\boldsymbol{z}, \phi), \tag{5.4}$$

where $a_0, a_1, a_2 \in \mathbb{R}$, $\mathbf{B}_0$ is any symmetric $n \times n$ matrix, and g_o is any function such that $g_o(\boldsymbol{z}, \phi) = -g_o(-\boldsymbol{z}, \phi)$ or $g_o(\boldsymbol{z}, \phi) = -g(\boldsymbol{z}, -\phi)$. It is claimed that this objective can be achieved with the quadrature node set

$$\mathcal{N} = \mathcal{N}^{00} \cup \mathcal{N}^{\eta 0} \cup \bigcup_{i=1}^{n} \mathcal{N}_i^{\xi 0} \tag{5.5}$$

Table 1
Constraints on the weights and nodes for the third-order quadrature method (5.6).

$f(\boldsymbol{z},\phi)$	Constraint
1	$w_{00}+2w_{\eta 0}+2nw_{\xi 0}=1$
$\cos\phi$	$w_{00}+2\cos\eta\, w_{\eta 0}+2nw_{\xi 0}=I_1(\kappa)/I_0(\kappa)$
$\cos 2\phi$	$w_{00}+2\cos 2\eta\, w_{\eta 0}+2nw_{\xi 0}=I_2(\kappa)/I_0(\kappa)$
${z_i}^2$, $i\in\{1,\dots,n\}$,	$2\xi^2 w_{\xi 0}=1$
${z_i}^4$, $i\in\{1,\dots,n\}$,	$2\xi^4 w_{\xi 0}=3$

for some choice of parameters ξ and η, and a quadrature rule of the form

$$\begin{aligned}&\int_{\mathbb{R}^n}\int_{-\pi}^{\pi}\mathcal{N}(\boldsymbol{z};\mathbf{0},\mathbf{I})\,\mathcal{VM}(\phi;0,\kappa)\,f(\boldsymbol{z},\phi)\,\mathrm{d}\phi\,\mathrm{d}\boldsymbol{z}\\ &\quad\approx w_{00}\sum_{(\boldsymbol{z},\phi)\in\mathcal{N}^{00}} f(\boldsymbol{z},\phi)+w_{\eta 0}\sum_{(\boldsymbol{z},\phi)\in\mathcal{N}^{\eta 0}} f(\boldsymbol{z},\phi)+w_{\xi 0}\sum_{i=1}^{n}\sum_{(\boldsymbol{z},\phi)\in\mathcal{N}_i^{\xi 0}} f(\boldsymbol{z},\phi).\end{aligned}\tag{5.6}$$

Indeed, to make (5.6) exact for functions of the form (5.4), it suffices to impose the condition that (5.6) be exact for the basis functions listed in Table 1. The corresponding constraints on the quadrature weights and the parameters ξ and η appearing in the quadrature nodes are also listed in the table. It is noted that these constraints follow from the assumed quadrature rule (5.6) in conjunction with the characteristic function (4.1) or the low-order moments listed in section 4.3. Solving the five constraint equations in Table 1 for w_{00}, $w_{\eta 0}$, $w_{\xi 0}$, ξ, and η yields

$$\begin{aligned}&\xi=\sqrt{3},\quad \eta=\arccos\left(\frac{B_2(\kappa)}{2B_1(\kappa)}-1\right),\\ &w_{\xi 0}=\frac{1}{6},\quad w_{\eta 0}=\frac{B_1(\kappa)^2}{4B_1(\kappa)-B_2(\kappa)},\quad w_{00}=1-2w_{\eta 0}-2nw_{\xi 0},\end{aligned}\tag{5.7}$$

where $B_p(\kappa)\equiv 1-I_p(\kappa)/I_0(\kappa)$.

In summary, the third-order GVM quadrature rule is (5.6) with the nodes and weights specified in (5.5) and (5.7). It is noted that the number of quadrature nodes in the set (5.5) is $|\mathcal{N}|=2n+3$. Thus, the number of nodes increases linearly with the dimension n. Further, this is precisely the same number of nodes (sigma points) used in the unscented transform (which is a third-order Gauss–Hermite quadrature method) in a Cartesian space of dimension $n+1$.

6. Uncertainty propagation. This section presents the method that implements the prediction step of the Bayesian nonlinear filter [12] using the GVM PDF as input. Effectively, the new algorithm provides a means for approximating the nonlinear transformation of a GVM distribution as another GVM distribution. The GVM quadrature rules developed in the previous section play a key role in the computation. In analogy to the unscented transform [13] applicable to Gaussian PDFs, sigma points (or quadrature nodes) are deterministically selected from the initial GVM distribution and then acted on by the nonlinear transformation. The transformed sigma points are then used to reconstruct the parameters of the transformed GVM

distribution. The main application of this methodology is the propagation of a space object's state uncertainty under nonlinear two-body dynamics, which provides improved prediction capabilities of the object's future location and characterization of its orbital uncertainty at future times.

The organization of this section is as follows. In subsection 6.1, the preliminary notation is defined followed by discussions on how state propagation under a nonlinear system of ordinary differential equations (ODEs) fits into the general framework. In subsection 6.2, the GVM filter prediction step is motivated followed by the complete algorithm description in subsection 6.3. In subsection 6.4, it is shown how the inclusion of additional uncertain parameters or stochastic process noise can be treated within the same framework.

6.1. Preliminary notation. Let $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$, $\boldsymbol{\Phi} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n \times \mathbb{S}$ be a diffeomorphism, and let $(\tilde{\boldsymbol{x}}, \tilde{\theta})$ be defined such that

$$(\tilde{\boldsymbol{x}}, \tilde{\theta}) = \boldsymbol{\Phi}(\boldsymbol{x}, \theta; \boldsymbol{p}), \tag{6.1}$$

where $\boldsymbol{p}$ denotes any constant nonstochastic parameters. The inverse of (6.1) is denoted as $(\boldsymbol{x}, \theta) = \boldsymbol{\Psi}(\tilde{\boldsymbol{x}}, \tilde{\theta}; \boldsymbol{p})$. Thus, given the GVM random vector $(\boldsymbol{x}, \theta)$ and a diffeomorphism $\boldsymbol{\Phi}$, the objective is to approximate the joint PDF of $(\tilde{\boldsymbol{x}}, \tilde{\theta})$ by a GVM distribution and quantify the fidelity of this approximation.

The first-order system of ODEs

$$\mathbf{x}'(t) = \mathbf{f}(\mathbf{x}(t), t) \tag{6.2}$$

naturally gives rise to a family of diffeomorphisms. For a given initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$, we denote the solution of (6.2) as

$$\mathbf{x}(t) = \boldsymbol{\Phi}(\mathbf{x}_0; t_0, t), \tag{6.3}$$

which maps the state $\mathbf{x}_0$ at some initial epoch to the state at a future time. (Existence and uniqueness of solutions are assumed on the interval $[t_0, t]$.) In this context, $\boldsymbol{\Phi}$ is called the *solution flow* and is of the form (6.1), where the parameter vector $\boldsymbol{p}$ contains the initial and final times. The *inverse solution flow* is denoted as

$$\mathbf{x}_0 = \boldsymbol{\Psi}(\mathbf{x}(t); t_0, t),$$

which maps the state $\mathbf{x}(t)$ at time t to the state $\mathbf{x}_0$ at some past time t_0. If the initial condition $\mathbf{x}_0$ is *uncertain* and described by a PDF, then (6.3) dictates how this PDF is transformed to a future epoch. The dynamics governing the two-body problem in orbital mechanics (see section 2.1) can be cast in the first-order form (6.2), where $\mathbf{x}$ is either inertial Cartesian position-velocity coordinates or equinoctial orbital elements.

In what follows, the following notation and definitions are required. Given $(\boldsymbol{x}, \theta) \in \mathbb{R}^n \times \mathbb{S}$ and a diffeomorphism $\boldsymbol{\Phi} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n \times \mathbb{S}$,

- $\boldsymbol{\Phi}_{\boldsymbol{x}} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n$ is the $\boldsymbol{x}$ component of $\boldsymbol{\Phi}$ (i.e., components 1 to n of $\boldsymbol{\Phi}$);
- $\Phi_\theta : \mathbb{R}^n \times \mathbb{S} \to \mathbb{S}$ is the θ component of $\boldsymbol{\Phi}$ (i.e., component $n+1$ of $\boldsymbol{\Phi}$);
- $\partial_{\boldsymbol{x}} \boldsymbol{\Phi}_{\boldsymbol{x}} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^{n \times n}$ is the *Jacobian* of $\boldsymbol{\Phi}_{\boldsymbol{x}}$ with respect to $\boldsymbol{x}$; the (i, j)th component is the partial derivative of the ith component of $\boldsymbol{\Phi}_{\boldsymbol{x}}$ with respect to the jth component of $\boldsymbol{x}$;

- $\partial_{\boldsymbol{x}}\Phi_\theta : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n$ is the *gradient* of Φ_θ with respect to $\boldsymbol{x}$;
- $\partial^2_{\boldsymbol{x}}\Phi_\theta : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^{n\times n}$ is the *Hessian* of Φ_θ with respect to $\boldsymbol{x}$.

6.2. Motivation. Consider the random variables $(\tilde{\boldsymbol{x}}, \tilde{\theta})$ defined by

$$(\tilde{\boldsymbol{x}}, \tilde{\theta}) = \boldsymbol{\Phi}(\boldsymbol{x}, \theta; \boldsymbol{p}),$$

where $\boldsymbol{\Phi} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n \times \mathbb{S}$ is a diffeomorphism and $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$. By the change of variables theorem for PDFs, it follows that

$$p(\tilde{\boldsymbol{x}}, \tilde{\theta}) = \det\left[\frac{\partial \boldsymbol{\Psi}}{\partial \tilde{\mathbf{x}}}\right] \mathcal{GVM}\big(\boldsymbol{\Psi}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}}, \tilde{\theta}; \boldsymbol{p}), \Psi_\theta(\tilde{\boldsymbol{x}}, \tilde{\theta}; \boldsymbol{p}); \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa\big), \tag{6.4}$$

where $\boldsymbol{\Psi}$ is the inverse of $\boldsymbol{\Phi}$ and $\tilde{\mathbf{x}} = (\tilde{\boldsymbol{x}}, \tilde{\theta})$. The transformed PDF (6.4) is a GVM distribution in $(\tilde{\boldsymbol{x}}, \tilde{\theta})$ if the transformation $\boldsymbol{\Phi}$ (or $\boldsymbol{\Psi}$) is of the form (4.2). In general, the nonlinear transformation of a GVM distribution is not a GVM distribution. In such cases, the objective of the GVM filter prediction step is to compute an approximate GVM distribution such that

$$p(\tilde{\boldsymbol{x}}, \tilde{\theta}) \approx \mathcal{GVM}(\tilde{\boldsymbol{x}}, \tilde{\theta}; \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{P}}, \tilde{\alpha}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Gamma}}, \tilde{\kappa}). \tag{6.5}$$

One way to achieve this objective is to approximate $\boldsymbol{\Phi}$ as a Taylor series about the modal point $(\boldsymbol{x}, \theta) = (\boldsymbol{\mu}, \alpha)$ and then match the Taylor coefficients with the coefficients in the transformation (4.2). It follows that

$$\begin{aligned}
&\mathbf{L} = \partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha), \quad \boldsymbol{d} = \boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha) - \partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)\,\boldsymbol{\mu},\\
&a = \Phi_\theta(\boldsymbol{\mu}, \alpha) - \alpha - \partial_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)^T\boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{\mu}^T\partial^2_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)\,\boldsymbol{\mu},\\
&\boldsymbol{b} = \partial_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha) - \partial^2_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)\,\boldsymbol{\mu}, \quad \mathbf{C} = \partial^2_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha),
\end{aligned} \tag{6.6}$$

where, for notational convenience, the dependence of $\boldsymbol{\Phi}$ and its derivatives on the nonstochastic parameter $\boldsymbol{p}$ is suppressed. Substituting (6.6) into (4.3) yields

$$\begin{aligned}
&\tilde{\boldsymbol{\mu}} = \boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha), \quad \tilde{\mathbf{P}} = \partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)\,\mathbf{P}\,\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)^T, \quad \tilde{\alpha} = \Phi_\theta(\boldsymbol{\mu}, \alpha),\\
&\tilde{\boldsymbol{\beta}} = \mathbf{Q}^T\big[\boldsymbol{\beta} + \mathbf{A}^T\partial_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)\big], \quad \tilde{\boldsymbol{\Gamma}} = \mathbf{Q}^T\big[\boldsymbol{\Gamma} + \mathbf{A}^T\partial^2_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)\,\mathbf{A}\big]\mathbf{Q}, \quad \tilde{\kappa} = \kappa,
\end{aligned} \tag{6.7}$$

where $\mathbf{Q}$ is an $n \times n$ orthogonal matrix and $\mathbf{R}$ is an $n \times n$ positive-definite upper-triangular matrix such that $\mathbf{QR} = (\mathbf{LA})^T$. It follows that $\mathbf{Q} = \mathbf{A}^T\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)^T\tilde{\mathbf{A}}^{-T}$, where $\tilde{\mathbf{A}}$ is the lower-triangular matrix such that $\tilde{\mathbf{P}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$; hence

$$\begin{aligned}
&\tilde{\boldsymbol{\mu}} = \boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha), \quad \tilde{\mathbf{P}} = \partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)\,\mathbf{P}\,\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)^T, \quad \tilde{\alpha} = \Phi_\theta(\boldsymbol{\mu}, \alpha),\\
&\tilde{\boldsymbol{\beta}} = \tilde{\mathbf{A}}^{-1}\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)\,\mathbf{A}\big[\boldsymbol{\beta} + \mathbf{A}^T\partial_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)\big],\\
&\tilde{\boldsymbol{\Gamma}} = \tilde{\mathbf{A}}^{-1}\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)\,\mathbf{A}\big[\boldsymbol{\Gamma} + \mathbf{A}^T\partial^2_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu}, \alpha)\,\mathbf{A}\big]\mathbf{A}^T\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu}, \alpha)^T\tilde{\mathbf{A}}^{-T}, \quad \tilde{\kappa} = \kappa.
\end{aligned} \tag{6.8}$$

It is noted that the equations in (6.8) are analogous to those used in the prediction step of the EKF for the linearized propagation of a multivariate Gaussian distribution. Indeed, if $\boldsymbol{x} \sim$

$N(\boldsymbol{\mu}, \mathbf{P})$, $\boldsymbol{\Phi} : \mathbb{R}^n \to \mathbb{R}^n$, then, under weak nonlinear assumptions in $\boldsymbol{\Phi}$, $\tilde{\boldsymbol{x}} = \boldsymbol{\Phi}(\boldsymbol{x}) \dot{\sim} N(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{P}})$, where

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\Phi}(\boldsymbol{\mu}), \quad \tilde{\mathbf{P}} = \partial_{\boldsymbol{x}} \boldsymbol{\Phi}(\mu)\, \mathbf{P}\, \partial_{\boldsymbol{x}} \boldsymbol{\Phi}(\mu)^T.$$

While the analogous "EKF-like" prediction derived in (6.8) provides an approximate GVM distribution to a nonlinear transformation of a GVM random vector, the proposed algorithm developed in the next subsection avoids the direct computation of the partial derivatives of the transformation $\boldsymbol{\Phi}$.

6.3. Algorithm description. A complete description of the algorithm for transforming a GVM distribution under a nonlinear transformation and approximating the output as a GVM distribution is as follows. There are two inputs: (i) a random vector $(\boldsymbol{x}, \theta) \in \mathbb{R}^n \times \mathbb{S}$ distributed as a GVM distribution with parameter set $(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$, and (ii) a diffeomorphism $\boldsymbol{\Phi} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n \times \mathbb{S}$. The output is an approximation of the distribution of $(\tilde{\boldsymbol{x}}, \tilde{\theta}) = \boldsymbol{\Phi}(\boldsymbol{x}, \theta)$ as a GVM distribution with parameter set $(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{P}}, \tilde{\alpha}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Gamma}}, \tilde{\kappa})$. The algorithm has four main steps summarized below.

1. Generate N sigma points $\{(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i})\}_{i=1}^N$ from the input GVM distribution, and compute the transformed sigma points $(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}) = \boldsymbol{\Phi}(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i})$ for $i = 1, \ldots, N$.
2. Recover the parameters $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{P}}$ of the transformed GVM distribution from the transformed sigma points computed in Step 1 using the appropriate GVM quadrature rule.
3. Compute approximations of $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\Gamma}}$, denoted as $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Gamma}}$, using (6.8) in conjunction with suitable approximations of the partial derivatives of $\boldsymbol{\Phi}$.
4. Set $\tilde{\kappa} = \kappa$, and recover the parameters $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\Gamma}}$ by solving the least squares problem

$$(\tilde{\alpha}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Gamma}}) = \underset{\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Gamma}}}{\arg\min} \sum_{i=1}^{N} \left[M(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i}; \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa) - M(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}; \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{P}}, \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Gamma}}, \tilde{\kappa}) \right]^2, \tag{6.9}$$

where M is the Mahalanobis von Mises statistic (4.10).

These four steps are described in more detail below, including discussions on how they are specialized to the case when $\boldsymbol{\Phi}$ is the solution flow (in equinoctial orbital element coordinates) corresponding to the two-body problem in orbital mechanics (henceforth referred to as the "two-body problem").

6.3.1. Step 1. Using the GVM quadrature rule derived in section 5.1, first generate N quadrature nodes (sigma points) $\{(\boldsymbol{z}_{\sigma_i}, \phi_{\sigma_i})\}_{i=1}^N$ with respective weights $\{w_{\sigma_i}\}_{i=1}^N$ corresponding to the canonical GVM distribution. Next, use these canonical sigma points to generate sigma points corresponding to the input GVM random vector $(\boldsymbol{x}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ according to

$$\boldsymbol{x}_{\sigma_i} = \boldsymbol{\mu} + \mathbf{A}\boldsymbol{z}_{\sigma_i}, \quad \theta_{\sigma_i} = \phi_{\sigma_i} + \alpha + \boldsymbol{\beta}^T \boldsymbol{z}_{\sigma_i} + \tfrac{1}{2} \boldsymbol{z}_{\sigma_i}^T \boldsymbol{\Gamma} \boldsymbol{z}_{\sigma_i}$$

for $i = 1, \ldots, N$. Finally, compute the transformed sigma points $(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}) = \boldsymbol{\Phi}(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i})$ for $i = 1, \ldots, N$.

In the case of the two-body problem, the transformed sigma points are generated by propagating each sigma point $(\boldsymbol{x}_{\sigma_i}, \theta_{\sigma_i})$, representing a state in equinoctial orbital element space, from a specified initial time to a specified final time under the underlying dynamics. This generally requires the numerical solution of a nonlinear system of ODEs.

6.3.2. Step 2. The transformed parameters $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{P}}$ are given by

$$\tilde{\boldsymbol{\mu}} = \mathrm{E}[\tilde{\boldsymbol{x}}] = \mathrm{E}[\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{x},\theta)], \quad \tilde{\mathbf{P}} = \mathrm{E}\big[(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{\mu}})(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{\mu}})^T\big] = \mathrm{E}\big[(\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{x},\theta)-\tilde{\boldsymbol{\mu}})(\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{x},\theta)-\tilde{\boldsymbol{\mu}})^T\big].$$

Using the quadrature weights $\{w_{\sigma_i}\}_{i=1}^N$ and the transformed sigma points $\{(\tilde{\boldsymbol{x}}_{\sigma_i},\tilde{\theta}_{\sigma_i})\}_{i=1}^N$ generated in Step 1, the above expected values are approximated using the framework of the GVM quadrature developed in section 5:

$$\tilde{\boldsymbol{\mu}} = \sum_{i=1}^N w_{\sigma_i}\tilde{\boldsymbol{x}}_{\sigma_i}, \quad \tilde{\mathbf{P}} = \sum_{i=1}^N w_{\sigma_i}(\tilde{\boldsymbol{x}}_{\sigma_i}-\tilde{\boldsymbol{\mu}})(\tilde{\boldsymbol{x}}_{\sigma_i}-\tilde{\boldsymbol{\mu}})^T.$$

The lower-triangular Cholesky factor of $\tilde{\mathbf{P}}$, satisfying $\tilde{\mathbf{P}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$, is required in subsequent computations.

6.3.3. Step 3. The parameters $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\Gamma}}$ are approximated as $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Gamma}}$ using (6.8):

$$\begin{aligned}\hat{\alpha} &= \Phi_\theta(\boldsymbol{\mu},\alpha), \quad \hat{\boldsymbol{\beta}} = \tilde{\mathbf{A}}^{-1}\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu},\alpha)\,\mathbf{A}\big[\boldsymbol{\beta}+\mathbf{A}^T\partial_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu},\alpha)\big],\\ \hat{\boldsymbol{\Gamma}} &= \tilde{\mathbf{A}}^{-1}\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu},\alpha)\,\mathbf{A}\big[\boldsymbol{\Gamma}+\mathbf{A}^T\partial_{\boldsymbol{x}}^2\Phi_\theta(\boldsymbol{\mu},\alpha)\,\mathbf{A}\big]\mathbf{A}^T\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu},\alpha)^T\tilde{\mathbf{A}}^{-T}.\end{aligned} \tag{6.10}$$

It is noted that $(\boldsymbol{\mu},\alpha)$ is itself a sigma point corresponding to the canonical sigma point $(\boldsymbol{z},\phi) = (\mathbf{0},0)$; hence $\hat{\alpha}$ is trivially recovered. The partial derivatives of $\boldsymbol{\Phi}$ in the expressions for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Gamma}}$ can be computed directly. Alternatively, if the diffeomorphism $\boldsymbol{\Phi}$ is the solution flow which maps the initial state $(\boldsymbol{x},\theta)$ at time t_0 to a propagated state $(\tilde{\boldsymbol{x}},\tilde{\theta})$ at time t_f governed by the dynamics of the two-body problem in orbital mechanics, then $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Gamma}}$ can be approximated using unperturbed two-body dynamics. Indeed, if the state space describes equinoctial orbital elements [1] with $(\boldsymbol{x},\theta) = (a,h,k,p,q,\ell)$, then

$$\partial_{\boldsymbol{x}}\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\mu},\alpha) \approx \mathbf{I}, \;\; \partial_{\boldsymbol{x}}\Phi_\theta(\boldsymbol{\mu},\alpha) \approx \begin{bmatrix} -\frac{3}{2}n_0/a_0\,\Delta t \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \;\; \partial_{\boldsymbol{x}}^2\Phi_\theta(\boldsymbol{\mu},\alpha) \approx \begin{bmatrix} \frac{15}{4}n_0/a_0^2\,\Delta t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where $a_0 = \mu_1$, $n_0 = \sqrt{\mu_\oplus/a_0^3}$, $\mu_\oplus$ is the gravitational parameter, and $\Delta t = t_f - t_0$.

6.3.4. Step 4. The hatted parameters $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Gamma}}$ computed in Step 3 are only approximations to the "optimal" parameters $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\Gamma}}$ characterizing the output GVM distribution. In this step, the hatted parameters are "refined" by solving a least squares problem which is motivated as follows. Define

$$p_a(\tilde{\boldsymbol{x}},\tilde{\theta}) = \mathcal{GVM}(\tilde{\boldsymbol{x}},\tilde{\theta};\tilde{\boldsymbol{\mu}},\tilde{\mathbf{P}},\hat{\alpha},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Gamma}},\tilde{\kappa}), \tag{6.11}$$

$$p_e(\tilde{\boldsymbol{x}},\tilde{\theta}) = \mathcal{GVM}\big(\boldsymbol{\Psi}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}},\tilde{\theta};\boldsymbol{p}),\Psi_\theta(\tilde{\boldsymbol{x}},\tilde{\theta};\boldsymbol{p});\boldsymbol{\mu},\mathbf{P},\alpha,\boldsymbol{\beta},\boldsymbol{\Gamma},\kappa\big). \tag{6.12}$$

Note that p_a is the approximate GVM distribution of the output using the hatted parameters, while p_e is the "exact" PDF, as given by (6.4), but with the assumption that the determinant

factor is unity (i.e., $\boldsymbol{\Phi}$ is volume preserving). It is acknowledged that this is suboptimal (only if $\boldsymbol{\Phi}$ is not volume preserving); although, in the context of the two-body problem, it is observed that the largest deviations from unity in the determinant factor are on the order of 10^{-4} for scenarios with the strongest nonconservative forces.

In this refinement step, the approximations of the hatted parameters are improved by "matching" the approximate and exact PDFs in (6.11) and (6.12) at the N transformed quadrature nodes $\{(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i})\}_{i=1}^{N}$. One way to accomplish this objective is to study the overdetermined system of algebraic equations

$$p_a(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}) = p_e(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}), \quad i = 1, \ldots, N,$$

in $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Gamma}}$. Instead, the analogous equations in "log space" are used by defining

$$\ell_a(\tilde{\boldsymbol{x}}, \tilde{\theta}) = -2 \ln \left[\frac{p_a(\tilde{\boldsymbol{x}}, \tilde{\theta})}{p_a(\tilde{\boldsymbol{x}}_*, \tilde{\theta}_*)} \right], \quad \ell_e(\tilde{\boldsymbol{x}}, \tilde{\theta}) = -2 \ln \left[\frac{p_e(\tilde{\boldsymbol{x}}, \tilde{\theta})}{p_e(\tilde{\boldsymbol{x}}_*, \tilde{\theta}_*)} \right],$$

where $(\tilde{\boldsymbol{x}}_*, \tilde{\theta}_*)$ denotes the modal point. It follows that

$$\ell_a(\tilde{\boldsymbol{x}}, \tilde{\theta}) = M(\tilde{\boldsymbol{x}}, \tilde{\theta}; \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{P}}, \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Gamma}}, \tilde{\kappa}), \quad \ell_e(\tilde{\boldsymbol{x}}, \tilde{\theta}) = M\big(\boldsymbol{\Psi}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}}, \tilde{\theta}; \boldsymbol{p}), \Psi_\theta(\tilde{\boldsymbol{x}}, \tilde{\theta}; \boldsymbol{p}); \boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa\big),$$

where M is the Mahalanobis von Mises statistic (4.10). Now set $\tilde{\kappa} = \kappa$ (see (6.8)), and solve the overdetermined system of algebraic equations

$$\ell_a(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}) = \ell_e(\tilde{\boldsymbol{x}}_{\sigma_i}, \tilde{\theta}_{\sigma_i}), \quad i = 1, \ldots, N,$$

for $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\Gamma}}$, in the least squares sense. This leads to the optimization problem (6.9), which can be expressed in the equivalent form

$$(\tilde{\alpha}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Gamma}}) = \underset{\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Gamma}}}{\arg\min} \sum_{i=1}^{N} \Big[r_i(\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Gamma}}) \Big]^2, \tag{6.13}$$

where

$$r_i(\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Gamma}}) = \boldsymbol{z}_{\sigma_i}^T \boldsymbol{z}_{\sigma_i} - \tilde{\boldsymbol{z}}_{\sigma_i}^T \tilde{\boldsymbol{z}}_{\sigma_i} + 4\kappa \left(\sin^2 \tfrac{1}{2} \phi_{\sigma_i} - \sin^2 \tfrac{1}{2} \tilde{\phi}_{\sigma_i} \right),$$

and

$$\tilde{\boldsymbol{z}}_{\sigma_i} = \tilde{\mathbf{A}}^{-1} (\tilde{\boldsymbol{x}}_{\sigma_i} - \tilde{\boldsymbol{\mu}}), \quad \tilde{\phi}_{\sigma_i} = \tilde{\theta}_{\sigma_i} - \hat{\alpha} - \hat{\boldsymbol{\beta}}^T \tilde{\boldsymbol{z}}_{\sigma_i} - \tfrac{1}{2} \tilde{\boldsymbol{z}}_{\sigma_i}^T \hat{\boldsymbol{\Gamma}} \tilde{\boldsymbol{z}}_{\sigma_i}.$$

Methods for solving nonlinear least squares problems, such as Gauss–Newton, full Newton, and quasi-Newton updates, along with globalization methods such as line search and trust region methods including Levenberg–Marquardt [4], are efficient and mature and will not be discussed further here. It is noted that the hatted parameters computed in Step 3 can be used as a starting iteration.

For the two-body problem, it is proposed to only optimize over the parameters $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and the $(1,1)$-component of $\hat{\boldsymbol{\Gamma}}$. (The largest change in the initial parameter matrix $\boldsymbol{\Gamma}$ to the transformed matrix $\tilde{\boldsymbol{\Gamma}}$ is in the $(1,1)$-component.) To facilitate the solution of the nonlinear least

squares problem (6.13) with these specializations, it is useful to note the partial derivatives of the residuals r_i:

$$\frac{\partial r_i}{\partial \hat{\alpha}} = 2\kappa \sin\tilde{\phi}_{\sigma_i}, \quad \frac{\partial r_i}{\partial \hat{\boldsymbol{\beta}}} = 2\kappa \sin\tilde{\phi}_{\sigma_i}\, \tilde{\boldsymbol{z}}_{\sigma_i}, \quad \frac{\partial r_i}{\partial \hat{\Gamma}_{11}} = \kappa \sin\tilde{\phi}_{\sigma_i} \left[\left(\tilde{\boldsymbol{z}}_{\sigma_i}\right)_{11}\right]^2.$$

In closing, it is noted that the optimization problem (6.13) is numerically well conditioned and, for the case of the two-body problem, is inexpensive to solve relative to the propagation of the quadrature nodes.

6.4. Inclusion of stochastic process noise. The inclusion of uncertain parameters or stochastic process noise in the transformation of a GVM random vector can be treated within the algorithm of subsection 6.3 using state augmentation. Indeed, let $\boldsymbol{\Phi} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{S} \to \mathbb{R}^n \times \mathbb{S}$ be a family of diffeomorphisms in $\boldsymbol{w} \in \mathbb{R}^m$ defined such that

$$(\tilde{\boldsymbol{x}}, \tilde{\theta}) = \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{w}, \theta; \boldsymbol{p}),$$

where $\boldsymbol{p}$ denotes any constant nonstochastic parameters. It is assumed that $(\boldsymbol{y}, \theta) \sim GVM(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$, where $\boldsymbol{w}$ is a Gaussian random vector and $\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{w})$ denotes the augmented random vector. In many uncertainty propagation methods, it commonly assumed that the state vector $(\boldsymbol{x}, \theta)$ is independent of the process noise random vector $\boldsymbol{w}$ and that $\boldsymbol{w}$ is additive so that $\boldsymbol{\Phi}$ enjoys the form

$$\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{w}, \theta; \boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}, \theta; \boldsymbol{p}) + \boldsymbol{g}(\boldsymbol{x}, \theta; \boldsymbol{p})\, \boldsymbol{w}.$$

In this formulation, such assumptions on the additivity of the process noise and the independence of $(\boldsymbol{x}, \theta)$ with $\boldsymbol{w}$ can be relaxed.

Introducing a slack variable $\tilde{\boldsymbol{w}}$, consider the augmented transformation

$$(\tilde{\boldsymbol{y}}, \tilde{\theta}) = \boldsymbol{\Omega}(\boldsymbol{y}, \theta; \boldsymbol{p}), \tag{6.14}$$

where $\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{w}})$ and

$$\boldsymbol{\Omega}(\boldsymbol{y}, \theta) = \begin{bmatrix} \boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{w}, \theta; \boldsymbol{p}) \\ \boldsymbol{w} \\ \Phi_\theta(\boldsymbol{x}, \boldsymbol{w}, \theta; \boldsymbol{p}) \end{bmatrix}.$$

It is noted that $\boldsymbol{\Omega}$ is a diffeomorphism from $\mathbb{R}^{n+m} \times \mathbb{S}$ to $\mathbb{R}^{n+m} \times \mathbb{S}$. Therefore, the algorithm of subsection 6.3 can be applied in conjunction with the transformation (6.14) to yield an approximation of the joint PDF of $(\tilde{\boldsymbol{y}}, \tilde{\theta})$ by a GVM distribution. Finally, to obtain the joint PDF of $(\tilde{\boldsymbol{x}}, \tilde{\theta})$, the slack variable $\tilde{\boldsymbol{w}}$ is marginalized out.

7. Results. In this section, an example is presented to demonstrate proof-of-concept of the GVM uncertainty propagation algorithm. The example is a scenario in low Earth orbit (LEO) which compares the GVM filter prediction step with that of the EKF [12], the UKF [13], a Gaussian sum filter (GSF) [9], and a particle filter [24]. The accuracy of the GVM uncertainty propagation algorithm is also validated using a metric based on the L_2 error. For the specific LEO scenario, it is shown that the GVM filter prediction step properly characterizes the actual uncertainty of a space object's orbital state while simple, less sophisticated, methods

which make Gaussian assumptions (such as the EKF and UKF) do not. Specifically, under the nonlinear propagation of two-body dynamics, the new algorithm properly characterizes the uncertainty for up to eight times as long as the standard UKF, all at no additional computational cost to the UKF.

In what follows, the particulars of the simulation scenario are defined in subsection 7.1, and the results are discussed in subsection 7.2.

7.1. Scenario description. This subsection describes the specific input to the uncertainty propagation algorithm in section 6.3, how the output is visualized, and how the accuracy of the output is validated.

7.1.1. Input. The initial GVM distribution of the space object's orbital state is defined with respect to equinoctial orbital elements $(a, h, k, p, q, \ell) \in \mathbb{R}^5 \times \mathbb{S}$ with parameter set $(\boldsymbol{\mu}, \mathbf{P}, \alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \kappa)$ given by

$$\boldsymbol{\mu} = (7136.635\,\mathrm{km}, 0, 0, 0, 0)^T, \quad \mathbf{P} = \mathrm{diag}\big((20\,\mathrm{km})^2, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}\big),$$
$$\alpha = 0, \quad \boldsymbol{\beta} = \mathbf{0}, \quad \boldsymbol{\Gamma} = \mathbf{0}, \quad \kappa = 3.282806 \times 10^7.$$

The mode of this distribution describes a circular, noninclined orbit in LEO with a semimajor axis of 7136.635 km. This choice of semimajor axis is made so that the instantaneous orbital period of the object is 100 minutes. In all subsequent discussions, a time unit of one orbital period is equal to 100 minutes. It is noted that the GVM distribution with the above parameter set is approximately Gaussian (since $\boldsymbol{\Gamma} = \mathbf{0}$ and $\kappa \gg 1$). In particular, the standard deviation in the mean longitude coordinate ℓ is $\sigma_\ell = 1/\sqrt{\kappa} = 0.01^\circ = 36''$. When validating against the EKF, UKF, or GSF, the osculating Gaussian (4.6) is used to convert the input GVM distribution to a Gaussian distribution. Finally, it is acknowledged that the initial parameter set defined above is representative of the uncertainty of a short two-minute radar track segment that has yet to be correlated to an existing object in the space catalog. Due to the sensitivity of releasing any real data acquired by the authors, this "sanitized example," motivated by the real data, is used instead.

The diffeomorphism describing the nonlinear transformation is the solution flow (6.3) induced from the system of ODEs (2.2) describing the two-body dynamics of orbital mechanics. (See the discussions in sections 2.1 and 6.1.) The parameters of the diffeomorphism are the epoch time t_0 of the input, the final epoch time t, and the specific forces used to model the perturbations. In all simulations, t_0 is the J2000 epoch (1 Jan 2000, 12:00 UTC) and $t - t_0$ is varied from 0.5 to 8 orbital periods. The EGM96 gravity model of degree and order 70 is used to model the perturbations. Though one can include additional nonconservative forces such as atmospheric drag, the nonlinearities induced by the gravity alone (especially in LEO) are sufficiently strong to stress the algorithm. Finally, the numerical integration of the ODEs is performed using a Gauss–Jackson method.

7.1.2. Visualization. The output of the uncertainty propagation algorithm in section 6.3 is a GVM distribution characterizing the uncertainty of the space object's orbital state at a specified future epoch. In order to visualize this six-dimensional PDF, the level curves (i.e., curves of equal likelihood) of the two-dimensional marginal PDF in the semimajor axis a and mean longitude ℓ coordinates are plotted. This choice is made because it is along this

particular two-dimensional slice where the greatest departure from "Gaussianity" and the most extreme banana- or boomerang-shaped level curves are observed.

In the panels of Figure 2, the $n\sigma$ level curves (of the marginal PDF in the a and ℓ coordinates) are plotted in half-sigma increments starting at $n = \frac{1}{2}$ for various final epoch times t. In order to visualize these marginal PDFs which are defined on a cylinder, the cylinder is cut at $\tilde{\alpha} - \pi$ (where $\tilde{\alpha}$ is the α parameter of the propagated GVM distribution) and rolled out to to form a two-dimensional plane. This plane is then rotated so that the semimajor and semiminor axes of the osculating Gaussian covariance are aligned with the horizontal and vertical. (Any such rotation or rescaling does not exaggerate any non-Gaussian effects or the extremity of the boomerang shape because all cumulants of order three and higher are invariant under an affine transformation.) These level curves are shown in shades of orange and red, as indicated. Where appropriate, overlays of the level curves generated from the EKF and UKF are shown in green and grey, respectively. Additionally, blue crosses represent particles generated from a Monte Carlo–based uncertainty propagator. If the represented PDF properly characterizes the actual uncertainty, then approximately 98.9% of the particles should be contained within the respective 3σ level curve.

7.1.3. Validation. An inspection of the panels in Figure 2 provides a simple visual means to assess whether the represented uncertainty properly characterizes the actual uncertainty of the space object's orbital state; "most" of the particles (blue crosses) should be contained within the level curves. To quantify uncertainty realism more rigorously and hence validate the prediction steps of the different filters under consideration, the normalized L_2 error is studied.

For functions $f, g : \mathcal{M} \to \mathbb{R}$, the *normalized* L_2 *error* between f and g is the scalar

$$\bar{L}_2(f, g) = \frac{\|f - g\|_{L_2}^2}{\|f\|_{L_2}^2 + \|g\|_{L_2}^2},$$

where $\|\cdot\|_{L_2}$ is the L_2 norm:

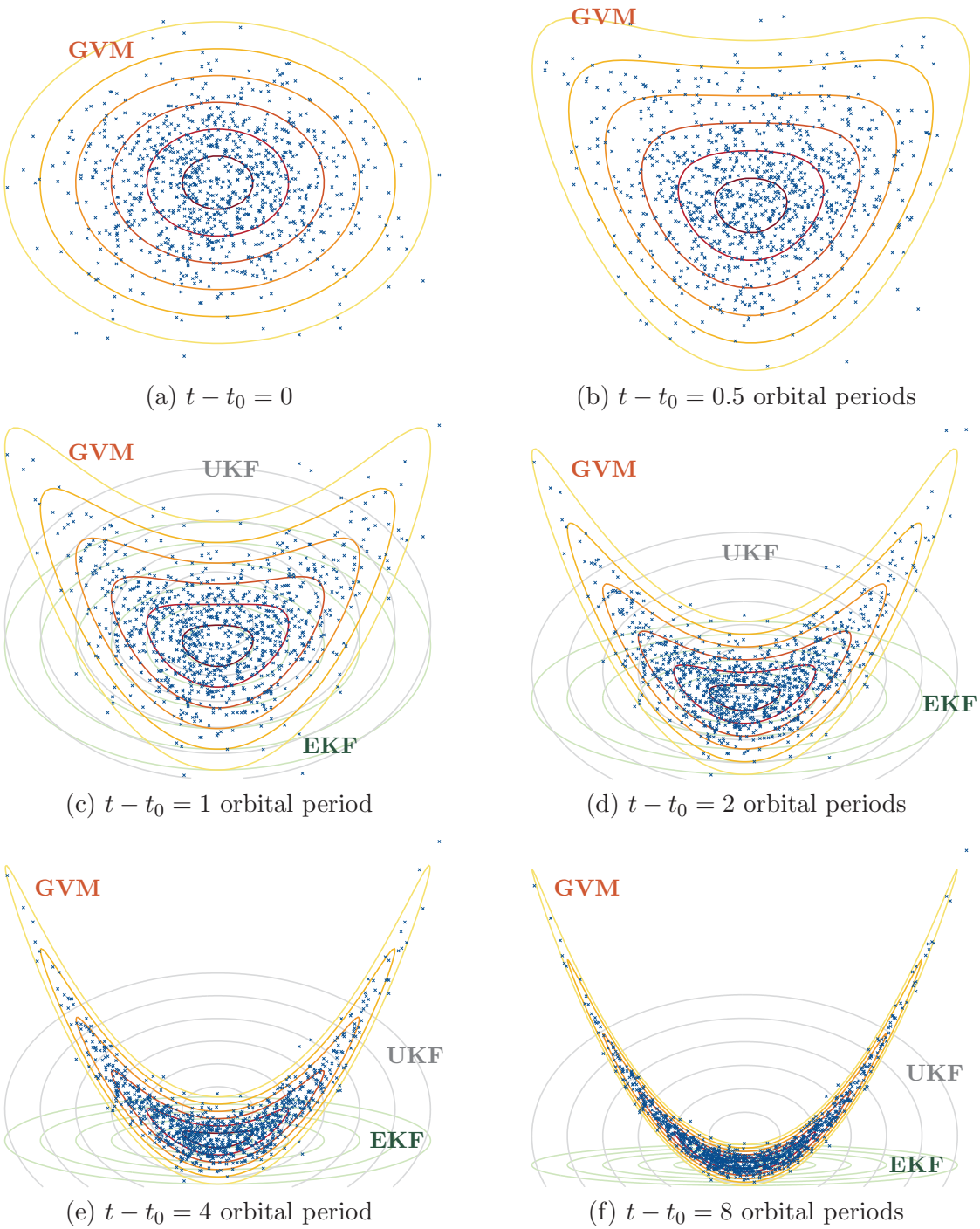
$$\|f\|_{L_2}^2 = \int_{\mathcal{M}} f(\boldsymbol{x})^2 \, \mathrm{d}\boldsymbol{x}.$$

By nonnegativity of the L_2 norm, $\bar{L}_2 \geq 0$ with equality if and only if $f = g$ in the L_2 sense. By the triangle inequality, $\bar{L}_2 \leq 1$ with equality if and only if f and g are orthogonal in the L_2 sense.

The validation tests shown in Figure 3 generate a time history of the normalized L_2 error

$$\bar{L}_2\big(p_{\text{approx}}(\cdot, t), p_{\text{baseline}}(\cdot, t)\big),$$

where t is the final epoch time. Further, $p_{\text{approx}}(\boldsymbol{u}, t)$ represents an approximation to the PDF of a space object's orbital state at time t ($\boldsymbol{u}$ represents equinoctial orbital elements) as computed by the prediction step of the GVM filter, EKF, UKF, or a GSF. Moreover, $p_{\text{baseline}}(\boldsymbol{u}, t)$ is a high-fidelity approximation to the exact state PDF which serves as a baseline for comparison. This baseline is computed using a high-fidelity Gaussian sum using the GSF of Horwood, Aragon, and Poore [9]. Thus, p_{approx} is a good approximation to the actual state PDF if the normalized L_2 error between it and the baseline p_{baseline} is zero or "close to zero" for all time.



(a) $t - t_0 = 0$

(b) $t - t_0 = 0.5$ orbital periods

(c) $t - t_0 = 1$ orbital period

(d) $t - t_0 = 2$ orbital periods

(e) $t - t_0 = 4$ orbital period

(f) $t - t_0 = 8$ orbital periods

Figure 2. *The uncertainty of a space object's orbital state at different epochs $t - t_0$ computed from the prediction steps of the EKF, UKF, GVM filter, and a particle filter. Shown are the respective level curves in the plane of the semimajor axis and mean longitude coordinates.*

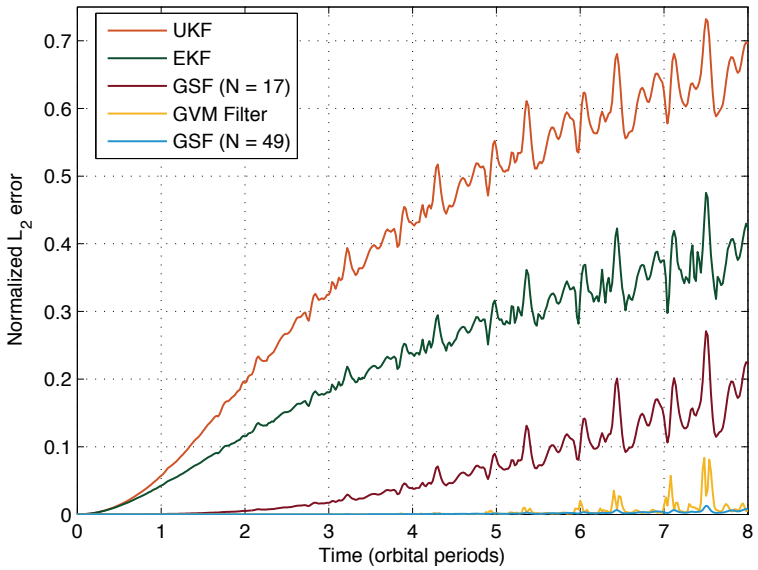


Figure 3. *Plots of the normalized L_2 error using different methods for uncertainty propagation.*

7.2. Discussions. As described in subsection 7.1.2, the panels in Figure 2 show the evolution of a space object's orbital uncertainty (with initial conditions defined in subsection 7.1.1) computed using the prediction steps of the EKF, UKF, GVM filter, and a particle filter. Each of the six panels shows the respective level curves in the plane of the semimajor axis and mean longitude coordinates at the epochs $t - t_0 = 0, 0.5, 1, 2, 4, 8$ orbital periods. In each of the six epochs, the level curves produced by the GVM filter prediction step (shown in shades of orange and red) correctly capture the actual uncertainty depicted by the particle ensemble. Each set of level curves is deduced from the propagation of only 13 sigma points (corresponding to a third-order GVM quadrature rule); the computational cost is the same as that of the UKF. For the UKF, its covariance (depicted by the grey ellipsoidal level curves) is indeed consistent (realistic) in the sense that it agrees with that computed from the definition of the covariance. Thus, in this scenario, the UKF provides "covariance realism" but clearly does not support "uncertainty realism" since the covariance does not represent the actual banana-shaped uncertainty of the exact PDF. Furthermore, the state estimate produced from the UKF coincides with the mean of the true PDF; however, the mean is displaced from the mode. Consequently, the probability that the object is within a small neighborhood centered at the UKF state estimate (mean) is essentially zero. The EKF, on the other hand, provides a state estimate coinciding closely with the mode, but the covariance tends to collapse, making inflation necessary to begin to cover the uncertainty. In neither the EKF nor the UKF case does the covariance actually model the uncertainty. In summary, *the GVM filter prediction step maintains a proper characterization of the uncertainty; the EKF and UKF do not.*

Figure 3 shows the evolution of the normalized L_2 error, as defined in subsection 7.1.3, computed from the (prediction steps of the) UKF, EKF, the GSF of Horwood, Aragon, and Poore [9] with $N = 17$ and $N = 49$ components, and the GVM filter. Uncertainty propagation

using the UKF and EKF quickly break down, but accuracy can be improved by increasing the fidelity of the Gaussian sum. The normalized L_2 errors produced from the GVM filter prediction step lie between those produced from the 17 and 49-term Gaussian sum. It is noted that the 17-term Gaussian sum requires the propagation of $17 \times 13 = 221$ sigma points; the GVM distribution requires only 13. If one deems a normalized L_2 error of $\bar{L}_2 = 0.05$ to signal a breakdown in accuracy, then the UKF and EKF prediction steps first hit this threshold after about one orbital period. By examining when the normalized L_2 error first crosses 0.05 for the GVM filter prediction, it is seen that *the GVM filter prediction step can faithfully propagate the uncertainty for about* 8 *times longer than when using an EKF or UKF.*

These results also suggest what could be achieved if the orbital state uncertainty is represented as a *mixture* of GVM distributions, in analogy to the Gaussian sum (mixture). As a single GVM distribution achieves accuracy (in the L_2 sense) between the 17- and 49-term Gaussian mixture, it is hypothesized that a GVM mixture would require about 95% fewer components than a Gaussian mixture to achieve comparable accuracy and uncertainty realism. Thus, one can extend the validity over which the uncertainty is propagated without the ballooning cost of doing so. Extending the work of this paper to GVM mixtures is future research.

8. Conclusions. Space surveillance is one example of a tracking environment requiring both covariance realism and uncertainty realism with the latter necessitating a proper characterization of the full probability density function (PDF) of a space object's state. This paper provided the initial motivation, mathematical background, and definition of a new class of multivariate PDFs, called the Gauss von Mises (GVM) distribution, which provides a statistically rigorous treatment of a space object's uncertainty in an orbital element space. Using the new class of GVM PDFs within the general Bayesian nonlinear filter, the resulting filter prediction step (i.e., uncertainty propagation) was shown to be tractable and operationally viable by providing uncertainty realism at the same computational cost as the traditional UKF. Furthermore, the GVM filter prediction step was shown to maintain a proper characterization of the uncertainty for up to eight times as long as the UKF. The ability to properly characterize and propagate uncertainty, as demonstrated in this paper, is a prerequisite to many functions in space situational awareness (SSA) such as tracking, space catalog maintenance, and conjunction analysis.

Acknowledgments. The authors thank Jeff Aristoff and Navraj Singh for fruitful discussions.

REFERENCES

[1] R. A. Broucke and P. J. Cefola, *On the equinoctial orbit elements*, Celestial Mech. Dynam. Astronom., 5 (1972), pp. 303–310.

[2] K. J. DeMars, M. K. Jah, Y. Cheng, and R. H. Bishop, *Methods for splitting Gaussian distributions and applications within the AEGIS filter*, in Proceedings of the 22nd AAS/AIAA Space Flight Mechanics Meeting, Charleston, SC, 2012, AAS-12-261.

[3] O. E. Drummond, T. L. Ogle, and S. Waugh, *Metrics for evaluating track covariance consistency*, in Signal and Data Processing of Small Targets 2007, Proc. SPIE 6699, SPIE, Bellingham, WA, 2007, 669916.

[4] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, UK, 1991.
[5] K. Fujimoto and D. J. Scheeres, *Non-linear propagation of uncertainty with non-conservative effects*, in Proceedings of the 22nd AAS/AIAA Space Flight Mechanics Meeting, Charleston, SC, 2012, AAS-12-263.
[6] A. Genz and B. D. Keister, *Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight*, J. Comput. Appl. Math., 71 (1996), pp. 299–309.
[7] T. Gerstner and M. Griebel, *Numerical integration using sparse grids*, Numer. Algorithms, 18 (1998), pp. 209–232.
[8] G. Golub and C. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, MD, 1996.
[9] J. T. Horwood, N. D. Aragon, and A. B. Poore, *Gaussian sum filters for space surveillance: Theory and simulations*, J. Guidance Control Dynam., 34 (2011), pp. 1839–1851.
[10] J. T. Horwood and A. B. Poore, *Adaptive Gaussian sum filters for space surveillance*, IEEE Trans. Automat. Control, 56 (2011), pp. 1777–1790.
[11] J. T. Horwood and A. B. Poore, *Orbital state uncertainty realism*, in Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, Wailea, HI, 2012, pp. 356–365.
[12] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Dover, New York, 1970.
[13] S. J. Julier, J. K. Uhlmann, and H. F. Durant-Whyte, *A new method for the nonlinear transformation of means and covariances in filters and estimators*, IEEE Trans. Automat. Control, 55 (2000), pp. 477–482.
[14] R. H. Lyddane, *Small eccentricities or inclinations in the Brouwer theory of the artificial satellite*, Astronom. J., 68 (1963), pp. 555–558.
[15] P. C. Mahalanobis, *On the generalised distance in statistics*, Proc. Natl. Inst. Sci. India, 2 (1936), pp. 49–55.
[16] K. V. Mardia, *Algorithm AS* 86: *The von Mises distribution function*, Appl. Statist., 24 (1975), pp. 268–272.
[17] K. V. Mardia and P. E. Jupp, *Directional Statistics*, John Wiley & Sons, New York, 2000.
[18] K. V. Mardia and T. W. Sutton, *A model for cylindrical variables with applications*, J. Roy. Statist. Soc. Ser. B, 40 (1978), pp. 229–233.
[19] O. Montenbruck and E. Gill, *Satellite Orbits: Models, Methods, and Applications*, Springer, Berlin, 2000.
[20] R. S. Park and D. J. Scheeres, *Nonlinear mapping of Gaussian statistics: Theory and applications to spacecraft trajectory design*, J. Guidance Control Dynam., 29 (2006), pp. 1367–1375.
[21] R. S. Park and D. J. Scheeres, *Nonlinear semi-analytic methods for trajectory estimation*, J. Guidance Control Dynam., 30 (2007), pp. 1668–1676.
[22] A. B. Poore, *Multidimensional assignment formulation of data association problems arising from multitarget tracking and multisensor data fusion*, Comput. Optim. Appl., 3 (1994), pp. 27–57.
[23] D. B. Reid, *An algorithm for tracking multiple targets*, IEEE Trans. Automat. Control, 24 (1997), pp. 843–854.
[24] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, Boston, MA, 2004.
[25] S. A. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Dokl. Akad. Nauk SSSR, 4 (1963), pp. 240–243.
[26] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, *Uncertainty propagation for nonlinear dynamic systems using Gaussian mixture models*, J. Guidance Control Dynam., 31 (2008), pp. 1623–1633.
[27] M. J. H. Walker, B. Ireland, and J. Owens, *A set of modified equinoctial orbit elements*, Celestial Mech. Dynam. Astronom., 36 (1985), pp. 409–419.