01 Apr 2022

Information-Based Particle Flow with Convergence Control

Kari C. Ward

Kyle J. Demars
Missouri University of Science and Technology, demarsk@mst.edu

Information-Based Particle Flow With Convergence Control

KARI C. WARD [ID] , Graduate Student Member, IEEE
Missouri University of Science and Technology, MO, USA

KYLE J. DEMARS [ID] , Member, IEEE
Texas A&M University, TX, USA

A new formulation of the Gaussian particle flow filter is presented using an information theoretic approach. The developed information-based form advances the Gaussian particle flow framework in two ways: it imparts physical meaning to the flow dynamics and provides the ability to easily include modifications for a non-Bayesian update. An orbit determination simulation with high initial uncertainty is used to demonstrate the consistent, robust performance of the information flow filter in situations where the extended Kalman filter fails.

Authors' addresses: Kari C. Ward is with the Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO 65409 USA, E-mail: (kcwd45@mst.edu); Kyle J. DeMars is with the Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843 USA, E-mail: (demars@tamu.edu). (Corresponding author: Kari Ward.)

I. INTRODUCTION

Across a wide array of disciplines, Bayesian estimation is the foundation for incorporating new information into the understanding of a given stochastic process. The methods by which this update is performed vary widely in complexity and accuracy. For a linear, Gaussian process, the Kalman filter provides an exact, closed-form solution [1]. In non-linear estimation, the extended Kalman filter (EKF) uses a local linear approximation to update the state estimate through the same Kalman filter framework. Problems with this approximation arise in cases where the process is not well-modeled by either a Gaussian or local linear assumption. In practice, straying from this assumption degrades estimation accuracy, performance, and can result in filter divergence [1].

Significant research has been conducted on methods to alleviate nonlinear estimation issues through techniques such as Gaussian mixture modeling [2] or the particle filter [3], [4]. These approaches have been used extensively in nonlinear estimation problems, such as multitarget tracking [5], [6] and orbit determination [7]. By representing the probability density function (PDF) with a sufficiently large number of samples, particle filters propagate and update a set of particles and their corresponding weights to provide PDF moments. In applications where drawing samples directly from the density is impossible, importance sampling from a given proposal density can be used [4]. However, traditional particle filtering methods suffer from what is commonly referred to as the "curse of dimensionality." As the dimension of the state increases, the number of particles required to effectively approximate the distribution increases. For low state dimensions, the rate has been shown to be two orders of magnitude for each additional state dimension [8], and even in the linear case, the error in the approximation can grow exponentially [9].

The heart of the dimensionality problem lies in the concept of particle degeneracy [4]. In applying Bayes' rule, the prior estimate is updated to the posterior through the measurement likelihood. While a set of particles may adequately represent the prior distribution, not all of the particles are encompassed in the measurement likelihood. In the event of high precision measurements relative to the state uncertainty, the number of particles with a meaningful contribution in approximating the posterior distribution can be a drastically smaller subset of the total [9]. While the prior is sufficiently represented, the majority of the particles have little to no contribution in representing the posterior, i.e., their weights degenerate to zero. The instinctive solution to this problem is to use more particles to represent the prior in order to have an adequate number encompassed by the evidence. However, the computational burden is compounded by higher dimensions, larger uncertainties, and more precise measurements, all of which necessitate more particles to accurately represent the distribution.

To address problems in the filter performance, a method called particle flow has been developed for computing a Bayesian update [10], [11]. A set of particles is used to

represent the prior distribution, and the particles are moved through the state space to represent the posterior distribution according to Bayes' rule. In contrast to traditional particle filters where corresponding weights are updated based on new measurements and particle locations in the state space are left unchanged, particle flow equally and invariably weights the particles, opting instead to update their locations. By foregoing the weight update, particles undergoing particle flow need not be periodically resampled. The flow dynamics are built upon the evaluation of Bayes' rule formulated by a differential equation rather than the multiplication of two functions [11]. This probability "flow" from prior to posterior is governed by a logarithmic-homotopy that allows for the linear combination of prior and likelihood components and is applied in the state space to "map" between distributions. Through a time-like homotopy parameter that varies from zero to one, the prior distribution (at "time" zero) is smoothly morphed into the posterior distribution (at "time" one). Due to the underdetermined nature of the system, a multitude of flow solutions exist to describe the motion of the particles through the update. The most commonly implemented solution [11] uses a Gaussian assumption to define the flow in terms of two parameters that can be calculated from quantities familiar to an EKF update.

Information theoretic research has shown that Bayes' rule can alternatively be interpreted from an information perspective through a simultaneous Kullback–Leibler (KL) divergence minimization and expected logarithmic-likelihood maximization [12]. Other information/estimation connections have been established for additive Gaussian noise channels [13], as well as a bridge between information and estimation through the Fokker–Planck equation [14], [15]. In estimation problems where little-to-no initial knowledge of the state is available, the uncertainty can be difficult to handle with traditional filtering techniques, causing unstable behavior and poor estimation accuracy. An advantageous filter design for high uncertainty applications must promote robust performance in problems where little-to-no initial information is available or when such data is sparse. To that end, this article poses the particle flow framework from an information theoretic perspective to motivate the movement of particles through a Bayesian update based on the flow of new information.

The information theoretic approach established in this work seeks to expand the Gaussian particle flow filtering capabilities by addressing two avenues for improvement. First, the flow parameters governing the motion of the particles have no clear link to a physical or abstract process acting on the set of particles. That is, the method for deriving the flow parameters provides no immediately apparent understanding about the underlying motion of an arbitrary particle. Characterizing the particle motion illuminates the second contribution of this article, the inclusion of a non-Bayesian update. Several flow solutions exist that are capable of including something akin to noise or uncertainty in the particle flow update through mechanisms such as

Gromov's method for linear underdetermined partial differential equations [16] or geodesic manifolds [17]. However, these solutions are often achieved by an added complexity that negates the computational benefits of particle flow in the first place.

By circumventing the curse of dimensionality common to other particle filter formulations, particle flows are capable of providing equivalent or better filter performance with lower computational overhead [11]. However, such savings are highly dependent on the underlying flow design. Compared to the original Gaussian particle flow approach, even the local exact variation results in a 20-fold increase in computation time on a per-recursion basis [18]. For a linear, Gaussian sensor network example, stochastic particle flows have been shown to require computation times two orders of magnitude higher compared to more traditional particle filters [19]. Flows designed around geodesic manifolds have been reported to require 30 and 75 times more computations per particle than sequential importance resampling methods [17]. These flow designs trade a significant increase in computation time for improved performance.

This article approaches the problem from a different angle, by reformulating an existing and deterministic flow design to provide more control over the filter performance. The new particle motion model retains the Bayesian framework and is mathematically equivalent to its predecessor. However, the new model also provides unique opportunities to improve the filter performance using quantities defined for the flow equations. The new design retains the computational efficiency of the original Gaussian flow, and the novel modifications allow the filter to provide desirable performance without necessarily requiring more complex methods such as stochastic flows.

For challenging estimation problems, such as orbit determination with high initial state uncertainty or large periods of time between measurements, consistent, and accurate performance of the EKF is not guaranteed. A robust approach to the problem would involve computing a constrained admissible region of captured orbits and construction of a Gaussian mixture model [7] to initialize an appropriate filter. Combining the robust and efficient performance of particle flow filters (PFFs) with the ability to perform non-Bayesian updates allows the information flow filter (IFF) to serve as a middle ground between the EKF and methods like Gaussian mixture-type filters.

The rest of this article is organized as follows. First, a brief derivation of Gaussian particle flow is given in Section II, followed by the development of the proposed information flow in Section III. Methods for modifying the information flow update through entropic convergence control are provided in Section IV. Section V compares the EKF and information flow in a high uncertainty simulation to demonstrate the capabilities of the method with and without the developed convergence control. Finally, Section VI concludes this article.

II. PARTICLE FLOW

The fundamental purpose of the particle flow approach is to compute Bayes' rule through the solution of a differential equation. The uncertainty distribution is first represented using a set of particles to provide estimates of the PDF moments. The set can be drawn from an initial PDF or sampled by other means from the state space. Similar to the EKF, each particle is propagated in time through the state dynamics. For the update stage, particle flow assumes the motion of the particles from the prior to the posterior distribution follows an ordinary differential equation (ODE) that, in conjunction with a homotopy of Bayes' rule, can be used to generate solutions for the movement of said particles. The result is a continuous shift of the particles through the state space from locations according to the prior to new areas such that the posterior distribution is represented. The formulation of the logarithmic-homotopy, ODE, and subsequent Gaussian particle flow solution [11] is as follows.

Consider Bayes' rule

$$p(\boldsymbol{x}) = \frac{g(\boldsymbol{x})\ell(z|\boldsymbol{x})}{C}$$

where $p(\boldsymbol{x})$ is the posterior, $g(\boldsymbol{x})$ is the prior, $\ell(z|\boldsymbol{x})$ is the measurement likelihood, and C is the state-independent normalization constant necessary to ensure $p(\boldsymbol{x})$ is a valid PDF that integrates to one over the entire state space. Taking the logarithm of both sides yields

$$\log(p(\boldsymbol{x})) = \log(g(\boldsymbol{x})) + \log(\ell(z|\boldsymbol{x})) - \log(C). \quad (1)$$

Define a homotopy on (1) as

$$\log(\pi(\boldsymbol{x}(\lambda), \lambda)) = \log(g(\boldsymbol{x}(\lambda))) + \lambda \log(\ell(z|\boldsymbol{x}(\lambda))) \\ - \log(C(\lambda)) \quad (2)$$

such that, $\pi(\boldsymbol{x}(\lambda), 0)$ is the prior distribution and $\pi(\boldsymbol{x}(\lambda), 1)$ is the Bayesian posterior. Since Bayes' rule operates on an instantaneous moment in time, λ acts as a pseudotime quantity to propagate the homotopy distribution, $\pi(\boldsymbol{x}(\lambda), \lambda)$, from the prior to the posterior. As such, λ is referred to interchangeably as the homotopy parameter or pseudotime. It should be noted that $C(\lambda)$, the normalization term in (2), is functionally dependent on λ such that the homotopy distribution $\pi(\boldsymbol{x}(\lambda), \lambda)$ is a valid PDF for all λ.

Suppose that a particle $\boldsymbol{x}(\lambda)$ belonging to the PDF $\pi(\boldsymbol{x}(\lambda), \lambda)$ evolves in pseudotime according to the ODE

$$\frac{d\boldsymbol{x}(\lambda)}{d\lambda} = \boldsymbol{f}(\boldsymbol{x}(\lambda), \lambda).$$

The movement of the particle $\boldsymbol{x}(\lambda)$, with pseudotime functional dependence assumed hereafter, can be related to the change in its corresponding PDF via the Fokker–Planck equation; or, when assuming no stochastic diffusion, particle pseudotime evolution can be found via the Liouville equation [14]

$$\frac{\partial \pi(\boldsymbol{x}, \lambda)}{\partial \lambda} = -\nabla(\pi(\boldsymbol{x}, \lambda)\boldsymbol{f}(\boldsymbol{x}, \lambda)). \quad (3)$$

Expanding the gradient on the right-hand side of (3) results in

$$\frac{\partial \pi(\boldsymbol{x}, \lambda)}{\partial \lambda} = -\operatorname{tr}\left\{\frac{\partial \boldsymbol{f}(\boldsymbol{x}, \lambda)}{\partial \boldsymbol{x}}\right\}\pi(\boldsymbol{x}, \lambda) + \frac{\partial \pi(\boldsymbol{x}, \lambda)}{\partial \boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}, \lambda) \quad (4)$$

where $\operatorname{tr}\{\cdot\}$ denotes the trace operator. The partials of $\pi(\boldsymbol{x}, \lambda)$ with respect to λ and $\boldsymbol{x}$ are unknown but can be substituted for using the logarithmic-homotopy and properties of the PDF. Differentiating the logarithmic-homotopy in (2) with respect to λ and rearranging slightly gives

$$\frac{\partial \pi(\boldsymbol{x}, \lambda)}{\partial \lambda} = \pi(\boldsymbol{x}, \lambda)\left(\log(\ell(z|\boldsymbol{x})) - \frac{\partial C(\lambda)}{\partial \lambda}\right). \quad (5)$$

Assuming $\pi(\boldsymbol{x}, \lambda)$ is smooth and nowhere vanishing, the relationship

$$\frac{\partial \log(\pi(\boldsymbol{x}, \lambda))}{\partial \boldsymbol{x}} = \frac{1}{\pi(\boldsymbol{x}, \lambda)}\frac{\partial \pi(\boldsymbol{x}, \lambda)}{\partial \boldsymbol{x}} \quad (6)$$

can be used in (4) for the unknown partial of $\pi(\boldsymbol{x}, \lambda)$ with respect to $\boldsymbol{x}$. Utilizing the relationships given in (5) and (6) in (3), dividing by $\pi(\boldsymbol{x}, \lambda)$ and rearranging gives the general particle flow solution

$$\frac{\partial \log(\pi(\boldsymbol{x}, \lambda))}{\partial \boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}, \lambda) + \log(\ell(z|\boldsymbol{x})) - \frac{\partial C(\lambda)}{\partial \lambda} \\ = -\operatorname{tr}\left\{\frac{\partial \boldsymbol{f}(\boldsymbol{x}, \lambda)}{\partial \boldsymbol{x}}\right\}. \quad (7)$$

The system for the partial differential equation governing the particle flow in (7) is underdetermined and can be used to produce a multitude of solutions [20]. To arrive at the most commonly implemented solution, assume the flow function to be of the form

$$\boldsymbol{f}(\boldsymbol{x}, \lambda) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$$

and assume the prior and measurement likelihood are Gaussian densities. The Gaussian solution [11] can be found by substituting the Gaussian forms of the densities into (7). Denote the Gaussian PDF with mean $\boldsymbol{\theta}$ and covariance $\boldsymbol{\Theta}$ by $p_g(\cdot; \boldsymbol{\theta}, \boldsymbol{\Theta})$. Assuming $g(\boldsymbol{x}) = p_g(\boldsymbol{x}; \boldsymbol{m}^-, \boldsymbol{P}^-)$ and $\ell(z|\boldsymbol{x}) = p_g(z; \boldsymbol{H}\boldsymbol{x}, \boldsymbol{R})$ where $\boldsymbol{m}^-$ and $\boldsymbol{P}^-$ are the prior mean and covariance, z is the measurement, $\boldsymbol{H}$ is the mapping matrix from the state space to the measurement space such that the expected measurement is $\hat{z} = \boldsymbol{H}\boldsymbol{m}^-$, and $\boldsymbol{R}$ is the measurement noise covariance, (7) becomes

$$(-(\boldsymbol{x} - \boldsymbol{m}^-)^T \boldsymbol{P}^{-,-1}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) \\ + \lambda(z - \boldsymbol{H}\boldsymbol{x})^T \boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) \\ - \frac{1}{2}(z - \boldsymbol{H}\boldsymbol{x})^T \boldsymbol{R}^{-1}(z - \boldsymbol{H}\boldsymbol{x}) + \boldsymbol{c} = -\operatorname{tr}\{\boldsymbol{A}\} \quad (8)$$

where $\boldsymbol{c}$ denotes the collection of other terms not dependent on the state, $(\cdot)^{-1}$ is the matrix inverse, and $(\cdot)^T$ is the matrix transpose. The coefficients $\boldsymbol{A}$ and $\boldsymbol{b}$ can be solved for by equating terms that are quadratic and linear in $\boldsymbol{x}$; that is

$$\boldsymbol{x}^T \left(\boldsymbol{A}^T \boldsymbol{P}^{-,-1}\boldsymbol{m}^- - \boldsymbol{P}^{-,-1}\boldsymbol{b} + \lambda \boldsymbol{A}^T \boldsymbol{H}^T \boldsymbol{R}^{-1}z \\ -\lambda \boldsymbol{H}^T \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{b} + \boldsymbol{H}^T \boldsymbol{R}^{-1}z\right) = 0$$

$$x^T \left(-P^{-,-1}A - \lambda H^T R^{-1}HA - \frac{1}{2}H^T R^{-1}H \right) x = 0 .$$

Solving for the coefficients yields

$$A = -\frac{1}{2}P^- H^T \left(\lambda H P^- H^T + R \right)^{-1} H \qquad (9a)$$

$$b = (I + 2\lambda A) \left[(I + \lambda A)P^- H^T R^{-1}z + Am^- \right] \qquad (9b)$$

where I is an identity matrix of appropriate dimension.

In the event of a nonlinear measurement model, the expected measurement is assumed to be given by a nonlinear function $\hat{z} = h(m^-)$. The measurement, z, is replaced to reflect the nonlinearity by

$$z \leftarrow z - \hat{z} + Hm^- \qquad (10)$$

where H is then given by the Jacobian of the nonlinear measurement function with respect to the mean. Using the definition for z in (10) ensures that the term $(z - Hx)$ in (8) and subsequently, (9b), appropriately reflects a first-order Taylor series expansion about the prior mean to account for the nonlinear nature of the measurement function.

III. INFORMATION FLOW

For Gaussian particle flow, the parameters A and b in (9) are sufficient for modeling the movement of a particle, but their derivation does not provide a deeper understanding of the underlying processes driving that motion. Posing the problem from an information-theoretic perspective provides such an understanding by following the flow of new information from prior to posterior.

Incorporating new information or measurements into the filter framework involves updating the state distribution parameter estimates. For PFFs, this is accomplished by moving the particles from their *a priori* locations, through the state space, to new locations such that the set then represents the *a posteriori* belief. The dynamics of the information-theoretic particle flow (referred to herein as information flow) update are defined by a homotopy on the inclusion of information to the mean and covariance estimates in conjunction with an ODE governing the movement of individual particles [21].

Development of the underlying information homotopy begins with an information-theoretic equivalent to Bayes' rule [12]. It can be shown that the Bayesian posterior distribution is the one that simultaneously maximizes the expected measurement logarithmic-likelihood and minimizes the KL divergence between the prior and posterior, i.e.,

$$p(x) = \arg \min_{\pi(x)} \left\{ - \int \pi(x) \log(\ell(z|x)) dx \right.$$
$$\left. + \int \pi(x) \log \left(\frac{\pi(x)}{g(x)} \right) dx \right\} . \qquad (11)$$

If the prior, posterior, and measurement likelihood distributions are assumed to be Gaussian, the logarithmic-likelihood and KL divergence have closed-form, analytic solutions. Substituting the same Gaussian forms defined in Section II into (11) and assuming $\pi(x) = p_g(x; \mu, \Pi)$

allows a change of the optimization variable from the posterior distribution to its respective mean and covariance, or

$$\{m^+, P^+\} = \arg \min_{\mu, \Pi} \left\{ \frac{1}{2} \left[\log |P^- \Pi^{-1}| + \log |R| \right. \right.$$
$$+ (\mu - m^-)^T P^{-,-1}(\mu - m^-) + c + \mathrm{tr} \left\{ P^{-,-1}\Pi \right\}$$
$$\left. \left. + (z - H\mu)^T R^{-1}(z - H\mu) + \mathrm{tr} \left\{ H^T R^{-1}H\Pi \right\} \right] \right\} \qquad (12)$$

where $|\cdot|$ denotes the matrix determinant, m^+ and P^+ are the posterior mean and covariance, and c is all other terms not dependent on the state. Taking derivatives of the expression in the minimization of (12) with respect to μ and Π and solving for the parameters produces results similar in form and numerically equivalent to the Kalman filter update

$$\Pi = \left(P^{-,-1} + H^T R^{-1}H \right)^{-1}$$
$$\mu = \Pi \left(P^{-,-1}m^- + H^T R^{-1}z \right)$$

where the last term in both equations governs the incorporation of new information into the estimates. This division of prior- and information-based terms allows for the straightforward inclusion of a homotopy as

$$\Pi = \left(P^{-,-1} + \lambda H^T R^{-1}H \right)^{-1} \qquad (13a)$$
$$\mu = \Pi \left(P^{-,-1}m^- + \lambda H^T R^{-1}z \right) \qquad (13b)$$

such that when the homotopy pseudotime parameter λ is 0, μ and Π are equal to the prior mean and covariance, and when $\lambda = 1$, the standard form of the update, where μ and Π are the posterior mean and covariance, is recovered.

It should be noted that the solution for the homotopy mean and covariance in (13a) and (13b) can equivalently be determined by a homotopy on the expected logarithmic-likelihood in (11), resulting in

$$p(x) = \arg \min_{\pi(x)} \left\{ -\lambda \int \pi(x) \log(\ell(z|x)) dx \right.$$
$$\left. + \int \pi(x) \log \left(\frac{\pi(x)}{g(x)} \right) dx \right\} . \qquad (14)$$

Leveraging the Gaussian forms for the distributions in (14) to change the optimization variables to mean and covariance, setting the resulting function of Gaussian parameters to zero and solving results in the same solution for the homotopy distribution parameters as given in (13a) and (13b).

The motion of the individual particles during the update is taken to be governed by the ODE

$$\frac{dx}{d\lambda} = A(x - \mu) + \frac{d\mu}{d\lambda} \qquad (15)$$

where the terms are separated such that A and $\frac{d\mu}{d\lambda}$ govern the two modes of motion the particles undergo during the update. Specifically, contraction of the set relative to its center of mass (i.e., mean) and translation of the set through the state space. Defining the particle dynamics via the contraction and translation terms, respectively, ensures that the Gaussian covariance and mean are appropriately represented by the particles as λ goes from 0 to 1. The

pseudotime rate-of-change of the mean can be found by taking the derivative of (13b) with respect to λ, resulting in

$$\frac{d\boldsymbol{\mu}}{d\lambda} = \frac{d\boldsymbol{\Pi}}{d\lambda}\left(\boldsymbol{P}^{-,-1}\boldsymbol{m}^- + \lambda\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z}\right) + \boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} \quad (16)$$

where the pseudotime rate-of-change of the covariance is also defined by taking the derivative of (13a) with respect to λ

$$\frac{d\boldsymbol{\Pi}}{d\lambda} = -\boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{\Pi}\,. \quad (17)$$

The other flow parameter, $\boldsymbol{A}$ in (15), governs the relative contraction of the particle set in reference to its mean. As the uncertainty in the estimation is directly related to the entropy of the distribution, it is natural to define the change in the uncertainty, represented in information flow by the spread of the particles, using the change in entropy. The Shannon entropy [22], or in the case of a continuous distribution with PDF $q(\boldsymbol{x})$, the differential entropy, is defined in general as

$$H\left[q\right] = -\int q(\boldsymbol{x})\log q(\boldsymbol{x})d\boldsymbol{x}\,. \quad (18)$$

Following the approach of the work in [14], Liouville's equation can be used in (18) to relate the pseudotime rate-of-change of the homotopy distribution's entropy to the motion of a corresponding particle as

$$\frac{d}{d\lambda}H\left[\pi(\boldsymbol{x})\right] = \text{tr}\left\{\mathbb{E}\left\{\frac{d}{d\boldsymbol{x}}\left\{\frac{d\boldsymbol{x}}{d\lambda}\right\}\right\}\right\} \quad (19)$$

where $\mathbb{E}\{\cdot\}$ denotes the expected value. If the ODE in (15) is substituted into (19), the right-hand side can be simplified to

$$\frac{d}{d\lambda}H\left[\pi(\boldsymbol{x})\right] = \text{tr}\{\boldsymbol{A}\}\,. \quad (20)$$

If the Gaussian assumption is applied, the entropy can also be defined in terms of the covariance as

$$H\left[\pi(\boldsymbol{x})\right] = \frac{1}{2}\log|2\pi e\boldsymbol{\Pi}|$$

resulting in another definition for the pseudotime evolution of the entropy as

$$\frac{d}{d\lambda}H\left[\pi(\boldsymbol{x})\right] = \frac{1}{2}\text{tr}\left\{\boldsymbol{\Pi}^{-1}\frac{d\boldsymbol{\Pi}}{d\lambda}\right\}\,. \quad (21)$$

Using the covariance in (13a) and its corresponding pseudotime rate-of-change in (17) allows for the pseudotime rate-of-change of the entropy in (21) to be computed. Equating the two formulations for the pseudotime rate-of-change of entropy in (20) and (21) results in

$$-\frac{1}{2}\text{tr}\left\{\left(\boldsymbol{P}^{-,-1} + \lambda\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right\} = \text{tr}\{\boldsymbol{A}\}\,.$$

While the trace operation implies there are multiple valid solutions for $\boldsymbol{A}$, the chosen solution for $\boldsymbol{A}$ is given by equating the trace inner terms

$$\boldsymbol{A} = -\frac{1}{2}\left(\boldsymbol{P}^{-,-1} + \lambda\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H} \quad (22)$$

which is equivalent to the Gaussian flow $\boldsymbol{A}$ term given in (9a). Similarly, it can be shown that the $\boldsymbol{b}$ term given in

(9b) is equivalent to the remaining terms in (15) for the information flow, i.e.,

$$(\boldsymbol{I} + 2\lambda\boldsymbol{A})\left[(\boldsymbol{I} + \lambda\boldsymbol{A})\boldsymbol{P}^-\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} + \boldsymbol{A}\boldsymbol{m}^-\right]$$
$$= -\boldsymbol{A}\boldsymbol{\mu} + \frac{d\boldsymbol{\mu}}{d\lambda}\,.$$

Equivalency of the Gaussian particle flow and information flow solutions is demonstrated in a linear Gaussian example in [21]. Results of the simulation show that the differences between the state (position and velocity) estimates of the two Gaussian flows are unbiased and 16 orders of magnitude smaller than the states themselves; that is, approximately machine precision. It should also be noted that the information flow also makes use of the same modified measurement as the Gaussian particle flow [given in (10)] in the event of a nonlinear measurement model.

With the components and considerations for nonlinear models established, the process for the IFF, assuming nonlinear models, is as follows. First, for a given set of initial state mean and covariance estimates, the N_x particle set is sampled from the corresponding Gaussian density. Each sample is then propagated to the first measurement, at time k, using the dynamics model. At time step k, the prior mean and covariance are approximated by sample statistics of the particle set. The Jacobian $\boldsymbol{H}$ of the nonlinear measurement model is computed with respect to the prior mean, along with the modified measurement $\boldsymbol{z}$ following (10). Then, for each pseudotime step between 0 and 1, the homotopy mean and covariance are computed from (13), the mean pseudotime rate of change from (16) using the covariance rate of change in (17), and the entropy rate of change in (22). At this point, the ODE in (15) is integrated to move the particles and the cycle is repeated again at the next pseudotime step with the computation of the homotopy mean and covariance from (13). At $\lambda = 1$, the update is complete and the filter can begin again with the particle propagation to the next measurement time.

The calculation of the information flow components can also be simplified by incorporating a linear gain. From the information flow mean update in (13b), define the linear gain $\boldsymbol{K}_\lambda = \boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1}$. Recognizing that

$$\boldsymbol{K}_\lambda = \boldsymbol{\Pi}\boldsymbol{H}^T\boldsymbol{R}^{-1} = \boldsymbol{P}^-\boldsymbol{H}^T\left(\lambda\boldsymbol{H}\boldsymbol{P}^-\boldsymbol{H}^T + \boldsymbol{R}\right)^{-1} \quad (23)$$

and substituting into (13b) results in

$$\boldsymbol{\mu} = \boldsymbol{m}^- + \lambda\boldsymbol{K}_\lambda\boldsymbol{r}_z \quad (24)$$

where $\boldsymbol{r}_z = (\boldsymbol{z} - \hat{\boldsymbol{z}})$ is the measurement residual. The pseudotime rate-of-change of the mean can also be determined by substituting for the updated mean in (13b), the linear gain in (23), and the pseudotime rate-of-change of the updated covariance in (17), resulting in

$$\frac{d\boldsymbol{\mu}}{d\lambda} = \boldsymbol{K}_\lambda(\boldsymbol{r}_z - \lambda\boldsymbol{H}\boldsymbol{K}_\lambda\boldsymbol{r}_z)\,. \quad (25)$$

Finally, the contraction parameter in (22) can also be defined in terms of $\boldsymbol{K}_\lambda$ as

$$\boldsymbol{A} = -\frac{1}{2}\boldsymbol{K}_\lambda\boldsymbol{H}\,. \quad (26)$$

While the new information flow is equivalent to the Gaussian flow summarized in Section II, the new information-theoretic formulation provides a physical understanding of the motivating forces behind the movement of an arbitrary particle during a Bayesian update. By separating the flow parameters into contributions to translation and contraction of the set of particles, further examination and modification of the particle motion can be conducted.

IV. CONVERGENCE CONTROL

Estimation in situations with large state uncertainties and precise measurements can result in degraded performance, or in extreme cases, divergent filter behavior. When performing an update using a high precision measurement, a filter with little-to-no knowledge of the state tends to quickly converge on a solution that fits the limited information available. Often, the result of this uncertainty "snap-down" is a filter that is overconfident to some degree in its estimate; that is, the state estimation errors are no longer consistent with the filter's estimate of the corresponding uncertainty.

A primary source of the overconvergence issue lies in the linearization inherent to filters like the EKF and Gaussian-based particle flow methods. While the Bayesian solution provided by the information flow update is optimal, the magnitude of the demanded change in the estimate can violate the local linear assumption used to perform the update, resulting in poor estimation quality. To prevent such undesirable performance, filters like the EKF can utilize underweighting (UW) [23], through which the magnitude of the linear gain is reduced by some prescribed factor. There are a multitude of methodologies for indicating and performing an update with UW and examples of its use can be seen in both the Shuttle [24] and Orion [25] programs. One such approach, using the UW parameter, $0 < \beta < 1$, defines the conditions for UW as

$$||\boldsymbol{H}\boldsymbol{P}^-\boldsymbol{H}^T|| > \frac{1-\beta}{\beta}||\boldsymbol{R}|| \tag{27}$$

where $|| \cdot ||$ is the matrix L_2 norm. If the conditions in (27) are met, the gain reduction is applied through

$$\boldsymbol{K} = \boldsymbol{P}^-\boldsymbol{H}^T \left(\frac{1}{\beta}\boldsymbol{H}\boldsymbol{P}^-\boldsymbol{H}^T + \boldsymbol{R}\right)^{-1}. \tag{28}$$

An equivalent information flow update to that provided by the gain reduction in (28) can be produced by the exact same inclusion of β in the linear homotopy gain of (23). However, a direct incorporation of this or any other established appropriate UW technique does not leverage the unique mechanisms provided by the particle flow framework.

Separation of the information flow parameters according to their role in either translation or convergence of the set provides a convenient mechanism for implementing a novel modification to reduce overconvergence in the update. Consider the information flow parameter A in (22), defined by the pseudotime rate-of-change of the entropy. Introducing a new contraction efficiency, η, allows for direct control of the magnitude of the convergence [21]. The motion of a particle in information flow then becomes

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda} = \eta A(\boldsymbol{x} - \boldsymbol{\mu}) + \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}\lambda}. \tag{29}$$

Values of η less than one reduce the contraction of the particles while values greater than one increase the contraction for a particular pseudotime step. As the purpose of an entropic convergence control is explicitly to prevent overconvergence, values of η greater than one are not used, as they would only exacerbate the problem.

It should be noted that although the parameter A is identical in Gaussian particle and information flow, its contribution in the former is also tied to the other flow parameter, b in (9b), making it impossible to directly modify the particle flow convergence rate without first changing how the two parameters are defined. The ability to slow the uncertainty convergence during an update allows for better information flow filter performance by "softening" the update in situations with high uncertainty and precision measurements where the Bayesian update would in fact provide undesirable performance in subsequent time steps.

In the interest of furthering the robust filtering capabilities of information flow, a mathematically grounded basis for selecting the contraction efficiency parameter is presented. While η can be a constant, a more effective use is to define it as dependent on pseudotime and other flow-dependent quantities to provide variable convergence control when necessary and so as not to impede the flow otherwise. As the underlying purpose of convergence control is to lessen the aforementioned uncertainty snap-down, η can, therefore, most effectively be selected to control the pseudotime rate-of-change of the uncertainty, i.e., $\frac{\mathrm{d}\boldsymbol{\Pi}}{\mathrm{d}\lambda}$. While this term can be computed using the homotopy via (17), the realized pseudotime rate-of-change of the uncertainty is determined solely by the particle motion.

In [26], the temporal evolution of the entropy in a linearized system is used to detect nonlinear effects in the propagation of the uncertainty. Following that concept, the entropy, or more precisely its pseudotime evolution, can be used to detect particle contraction rates that produce undesirable, overconvergent behavior. One of the current problems in particle flow that is discussed is the need for a constant or approximately constant speed flow [27]. The flow ODE itself can exhibit severe stiffness, resulting in particle "velocities" at the beginning of the flow that can be several times larger than those toward the end of the flow. In the case of high uncertainties and precise measurements, the particle velocities at the beginning of the flow are further increased, resulting in potentially overconvergent behavior. Combining the findings of the work in [26] and the proposed capabilities of the work in [27], a convergence-rate constraint is applied through the entropy and contraction efficiency parameter.

Consider the two definitions for entropy evolution in (20) and (21), that is

$$\mathrm{tr}\{A\} = \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Pi}^{-1}\frac{\mathrm{d}\boldsymbol{\Pi}}{\mathrm{d}\lambda}\right\}. \tag{30}$$

As stated previously, although the covariance rate-of-change is ultimately the quantity that needs to be throttled, its realized rate-of-change is governed by the particles' contraction, or more specifically, A. This gives rise to two avenues to define a suitable contraction efficiency based on constant covariance (CC) rate-of-change and constant entropy (CE) rate-of-change.

To produce a CC rate-of-change threshold at the current pseudotime step, λ_n, values of Π and $\frac{d\Pi}{d\lambda}$ in at points λ_{CC} and λ_n, respectively, in the homotopy can be used in (30) to define the contraction efficiency parameter

$$\eta(\lambda_n) = \min \left\{ \frac{\text{tr}\{A_{\lambda_n}\}}{\frac{1}{2}\text{tr}\left\{\Pi_{\lambda_{CC}}^{-1}\left(\frac{d\Pi}{d\lambda}\right)_{\lambda_n}\right\}}, 1.0 \right\} \qquad (31)$$

where $\Pi_{\lambda_{CC}}^{-1}$ is given by evaluating (13a) at $\lambda = \lambda_{CC}$ and $(\frac{d\Pi}{d\lambda})_{\lambda_n}$ and A_{λ_n} AT $\lambda = \lambda_n$ in (17) and (26), respectively. For example, if $\lambda_{CC} = 1$, this defines A_{λ_n} as the required entropy rate-of-change to produce the posterior covariance (at $\lambda_{CC} = 1$) assuming a constant rate-of-change for the covariance equal to that of the current pseudotime step, i.e., $(\frac{d\Pi}{d\lambda})_{\lambda_n}$. The threshold for the convergence rate can be increased or decreased by selection of a different λ_{CC}. That is, using values of λ_{CC} less than one would allow for higher convergence rates while $\lambda_{CC} > 1$ would lower the maximum allowable converge rate. Selection of $\lambda_{CC} < 1$ can produce values of η greater than 1 for $\lambda_n > \lambda_{CC}$, necessitating the minimum check to constrain η on the interval $(0, 1]$.

Another, perhaps less abstract means of defining the contraction efficiency parameter is to directly limit the entropy rate-of-change at λ_n based on a CE rate-of-change at λ_{CE}. A limit based on this direct entropy rate comparison is given by

$$\eta(\lambda_n) = \min \left\{ \frac{\frac{1}{2}\text{tr}\left\{\Pi_{\lambda_{CE}}^{-1}\left(\frac{d\Pi}{d\lambda}\right)_{\lambda_{CE}}\right\}}{\text{tr}\{A_{\lambda_n}\}}, 1.0 \right\} \qquad (32)$$

such that when $\lambda_n < \lambda_{CE}$, $\eta A_{\lambda_n} = A_{\lambda_{CE}}$. The contraction efficiency formulation in (32) directly limits the rate of convergence at the beginning of the flow, where the entropy rates can be very high, to that of a later pseudotime step, resulting in slower convergence.

As is also the case with UW in an EKF application, the *a posteriori* mean and covariance estimates when using convergence control are not the same as their strictly Bayesian counterparts. However, the contraction efficiency parameter effectively constrains the particle velocities to prevent overconvergent behavior and produce more consistent velocities over a desired pseudotime interval. The result is a more conservative estimate, in terms of the uncertainty magnitude, compared to the results obtained without modifying the Bayesian approximate solution.

V. SIMULATION, ORBIT DETERMINATION

In order to demonstrate the high-uncertainty capabilities of the new IFF, an orbit determination simulation

TABLE I
Observation Schedule

Night	1	2	3	4	5	6	7	8	9	10	11
Start Meas.	1	64	127	169	N/A	N/A	190	206	269	353	395
Meas. Count	63	63	42	21	0	0	16	63	84	42	38

is used to compare performance with the EKF. Observations of a navigation-type satellite are simulated from the Maui Ground-based Electro-Optical Deep Space Surveillance (GEODSS) site over the course of eleven nights. A total of 432 optical (i.e., right ascension and declination) measurements are simulated with varying per-night totals, the details of which are given in Table I. The standard deviations of the measurement noises are 3 arcseconds each, and the initial state uncertainties (1σ) are 10 km and 0.25 m/s in position and velocity, respectively. The true initial state is taken as $[14525, -9478, 18714]$ km for the Earth-centered inertial frame x, y, and z positions, respectively, and $[3.248, 0.904, -2.062]$ km/s in the corresponding velocities. Two-body equations of motion [28] are used to describe the orbit dynamics. For each simulation, a total of 1000 trials is used to compute Monte Carlo (MC) statistics, with each trial's initial mean estimate drawn from a Gaussian distribution using the initial uncertainty and true initial state. A set of 1000 information flow initial particles is then drawn from a Gaussian distribution using the initial uncertainty and trial-specific mean estimate and is not resampled at any subsequent point during an individual trial.

To balance precision and computational burden in the information flow, discrete integration is used to perform the update with a pseudotime step of 0.01. Although the intuitive solution to integration of a stiff ODE such as the one used in the information flow update would be to use a stiff ODE solver, previous research [29] has indicated that discrete Euler integration with a sufficiently small step-size can, in general, produce desirable results with a significantly smaller computational burden compared to a stiff ODE solver.

A total of 12 simulations, 7 information flow, 2 EKFs, 2 Gaussian particle flows, and a single-trial extended Kalman particle filter (EKPF) are applied and compared to characterize the performance of the new filter and subsequent convergence control capabilities. The EKPF filter follows the approach outlined in [30]. In accordance with the findings of [31], a total of 10^6 particles are used in the EKPF approximation. The EKF simulations are comprised of a nominal EKF with no UW and one with UW following the approach outlined in (27) and (28) using $\beta = 5/6$. The PFFs follow the Gaussian exact (GEPFF) and Gaussian local exact (GLEPFF) formulations detailed in [18]. However, instead of using a parallel EKF to provide the prior and posterior covariances, the PFFs, like the IFFs, use the particle sample covariance. All other PFF simulation parameters (e.g., number of particles, flow integration) follow that of the nominal IFF. As stated in Section IV, the presented convergence control methods are not applicable to these Gaussian particle flows following the form of Section II,

Sim.	Label	Conv. Control
1	EKF Nominal	None
2	EKF $\beta = 5/6$	UW
3	IFF Nominal	None
4	IFF $\beta = 5/6$	UW
5	IFF $\lambda_{CC} = 1.2$	CC
6	IFF $\lambda_{CC} = 0.7$	CC
7	IFF $\lambda_{CC} = 0.5$	CC
8	IFF $\lambda_{CE} = 0.5$	CE
9	IFF $\lambda_{CE} = 0.2$	CE
10	GEPFF	None
11	GLEPFF	None
12	EKPF	None

due to the blending of contraction and translation related terms in the definition of $\boldsymbol{b}$ in (9b).

Table II gives the listing for the PFFs, EKPF, EKF, and IFFs simulations with UW, CC [see (31)] and CE [see (32)] rate-of-change convergence control selections. Values of λ_{CC} and λ_{CE} are chosen to develop an understanding of the IFF response to a wide range of valid selections for the parameters. Reflecting the CC definition, λ_{CC} selections are chosen in the neighborhood of 1. While $\lambda_{CC} = 1.2$ (Sim. 5) is not representative of a realizable pseudotime value in the flow, the homotopy covariance produced by that value is smaller, by definition, than the nominal final value, resulting in further increased throttling of the flow across the entire update. Due to the minimum check for both λ_{CC} and λ_{CE} in (31) and (32), respectively, both parameters can take on any values greater than zero. The selections chosen in Table II are intended to establish trends in performance relative to both the nominal and larger or smaller values of the parameters themselves in the neighborhood of reasonable choices for the parameters. To directly compare the filter behavior with both convergence control methods, a 0.5 case is examined for λ_{CC} and λ_{CE}.

To establish a baseline for comparison of the different simulations, the results for the nominal IFF are first examined. Average filter and MC error uncertainties (3σ) for the position states are given in Fig. 1 for the nominal IFF. Due to the large changes in uncertainty between different portions of the simulations, the left column of Fig. 1 gives results for the first 100 measurements while the right provides results for measurements 100 to 432. The results show several large spikes in uncertainty during the first night (measurement number <64) corresponding to the growth in uncertainty between observation arcs. Several smaller spikes in uncertainty can also be seen after the first night, again due to uncertainty growth between data collection times. As a result of the long gaps in time between some measurements, the results are given in reference to measurement time steps. Therefore, what is in fact an entire day of uncertainty growth looks like a near instantaneous spike when the scale changes. Results for the nominal IFF velocity states in Fig. 2 show similar instances of large spikes due to long propagation times between measurement sets followed by rapid convergence down to a steady state. Again, to
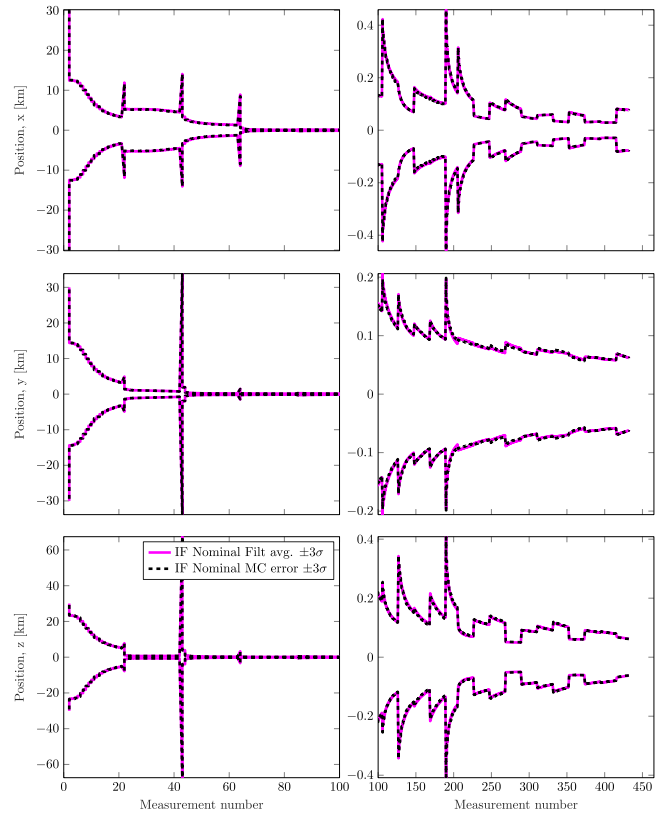

Fig. 1. Position results, filter average (solid) and MC error (dashed) $\pm 3\sigma$ for Sim. 3: nominal IFF. Left: measurements 1–100. Right: measurements 100–432.

alleviate the obscuring effects of large differences in uncertainty magnitude, the left column of Fig. 2 is given up to measurement 100 and the right column shows measurement 100 through 432. For both the position and velocity results in Figs. 1 and 2, the combination of large uncertainty reduction and similarity in the filter average and MC results makes it difficult to glean any other useful information from a direct $\pm 3\sigma$ comparison.

As noted in Section III, the information flow parameters are equivalent to those of the Gaussian flow presented in [11]. However, in contrast to the IFF, which only linearizes about the prior mean at $\lambda = 0$, the Gaussian PFFs included for comparison here linearize about the current particle set mean (GEPFF) or each particle (GLEPFF), for each λ. Differences in the MC error and filter average standard deviations for the GEPFF and GLEPFF with respect to the nominal IFF values are given in Fig. 3. By and large, the differences between the filters are small and without significant trends. The only exceptions being that the IFF has a slightly larger average filter standard deviation than the GLEPFF, and for the majority of the first two nights (measurement number <64), the IFF has a slightly smaller MC error standard deviation across the states. Otherwise, the three filters provide nearly indistinguishable performance with no filter outperforming the others over all 11 nights.

A more descriptive measure of the filter results, in terms of its consistency, is given in Figs. 4 and 5 for the position
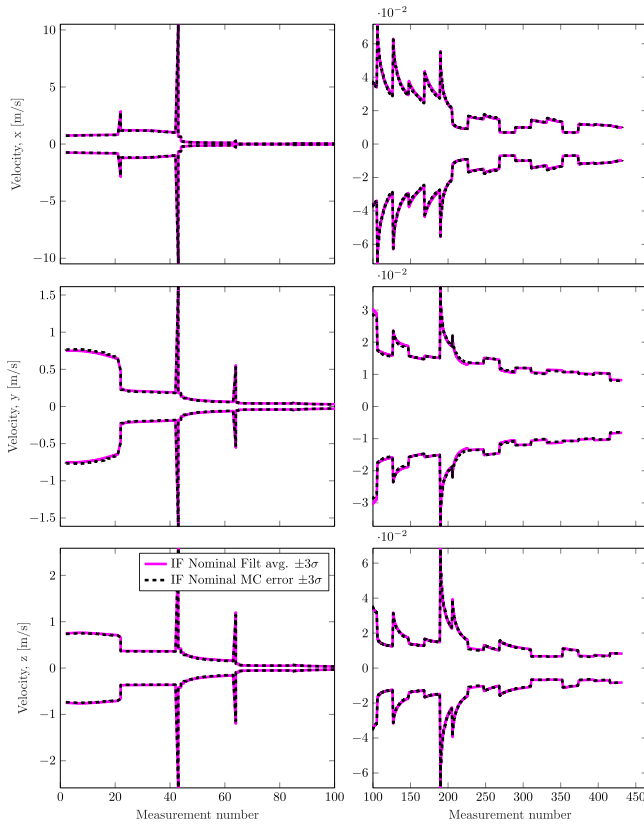
Fig. 2. Velocity results, filter average (solid) and MC error (dashed) $\pm 3\sigma$ for Sim. 3: nominal IFF. Left: measurements 1–100. Right: measurements 100–432.

and velocity states, respectively. The figures present three times the standard deviation for the ratio of single trial error to filter 3σ in each trial, for each simulation. For state j over the total N_t trials, the ratio at measurement ℓ is given by

$$\sigma_{\text{avg},j}^{\ell} = 3\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{e_{i,j}^{\ell}}{3\sigma_{i,j}^{\ell}} - \frac{1}{N_t}\sum_{k=1}^{N_t}\frac{e_{k,j}^{\ell}}{3\sigma_{k,j}^{\ell}}\right)^2} \quad (33)$$

where $e_{i,j}^{\ell}$ is the error at measurement ℓ in the trial i filter's posterior estimate for state j and $\sigma_{i,j}^{\ell}$ is the corresponding filter standard deviation. It should be noted that the 3 multipliers in (33) are simply included to emphasize the ratio is examining the relationship between the error distribution and the filter 3σ interval. That is, for each state, (33) gives the number of filter standard deviations, on average that would be needed to capture the MC error 3σ. For consistent filter performance, the value in (33) should be equal to or just slightly less than three. Values greater than three indicate overconfident or smug (i.e., overconvergent) filter performance while values less than three denote cautious, or conservative, filter performance. While neither drastically smug or cautious filter performance is desirable, ideally the filter would exhibit slightly cautious to consistent performance on average across all measurements.

The position results in Fig. 4 for the nominal IFF and two EKF simulations show that the IFF gives comparatively consistent performance with values in the neighborhood of 3.0,
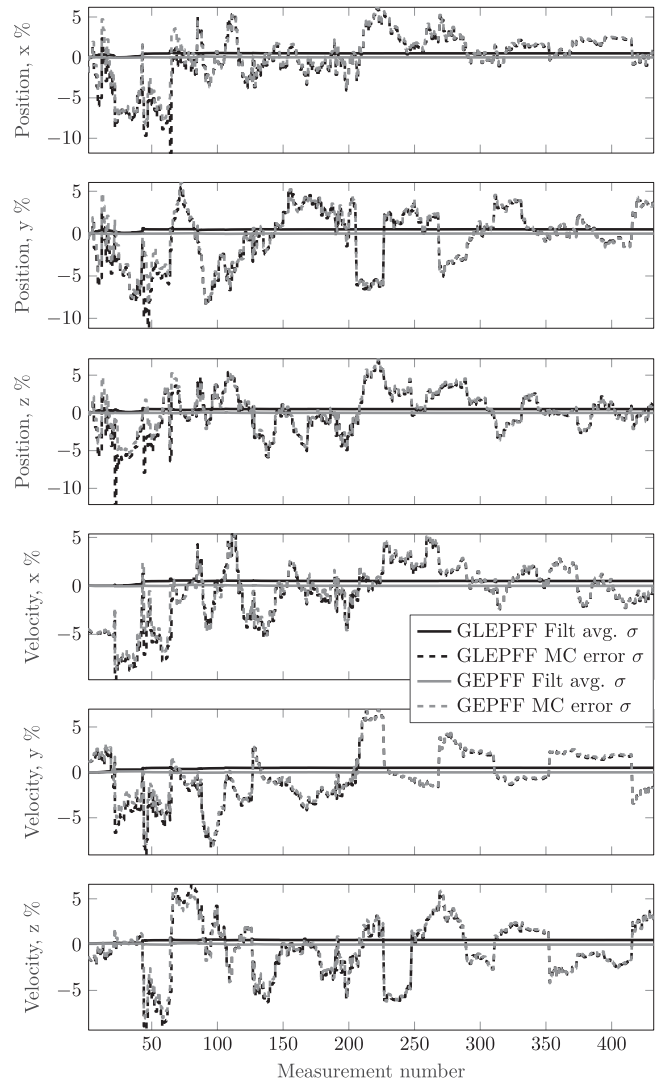

Fig. 3. GEPFF and GLEPFF PFFs: percent difference in MC error (dashed) and filter (solid) standard deviations with respect to the nominal IFF.

while the two EKFs show some cautious behavior before the second night of observations (measurement number <64) but then become successively more overconvergent through night 8 (measurement number <208). While the two EKFs exhibit a better filter performance in terms of consistency after night 8 (compared to nights 2–8), neither filter recovers a consistent or cautious estimate of the uncertainty across all states. As the velocity estimates are not directly related to the measurements and must be improved through correlations with the position states, the velocity results in Fig. 5 follow similar trends as their corresponding position channels. For all states, the EKF with UW shows slightly more consistent performance across the board compared to the EKF without, but the difference is small compared to the magnitude of the overconvergence exhibited by both EKFs. It should be noted that of the 1000 trials for the EKF simulations with and without UW, 112 and 92 trials failed, respectively. Failure in this context means the magnitude of a state error exceeded 10 times the filter estimated standard deviation for any
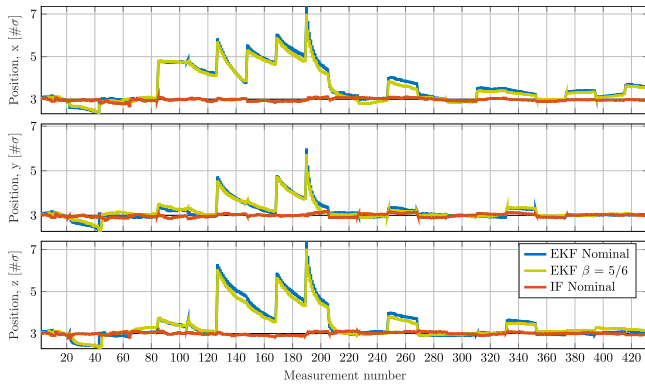
Fig. 4. Position results, average number of filter σ needed to capture MC error 3σ.
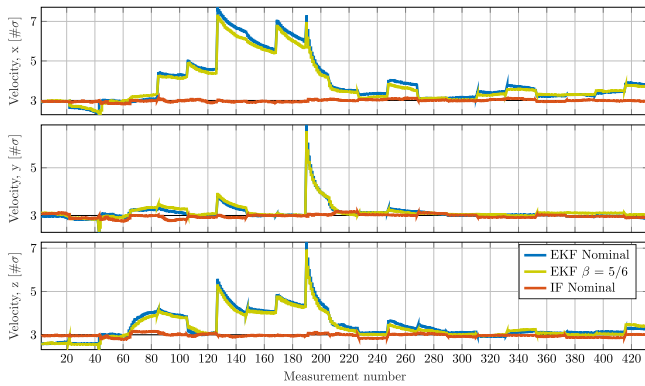

Fig. 5. Velocity results, average number of filter σ needed to capture MC error 3σ.
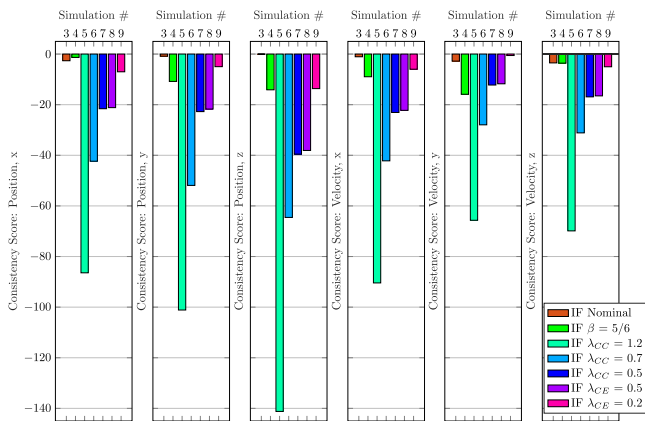
Fig. 6. Simulation scores for deviation from consistent estimation performance.

Fig. 7. Average position (top) and velocity (bottom) RSS values for each filter given in log-scale.

given state. Trials that are flagged as failures are removed from consideration in this analysis meaning the EKF MC statistics without and with UW are computed using 908 and 888 trials, respectively. It should also be noted that no trial for any IFF simulation met the failure criteria.

Results for the IFF simulations are given in Figs. 8 and 9 for the position and velocity σ ratios, respectively. As seen in the previous comparison with the EKFs in Figs. 4 and 5, the nominal IFF remains closely in the neighborhood of 3.0, straying slightly above or below depending on the
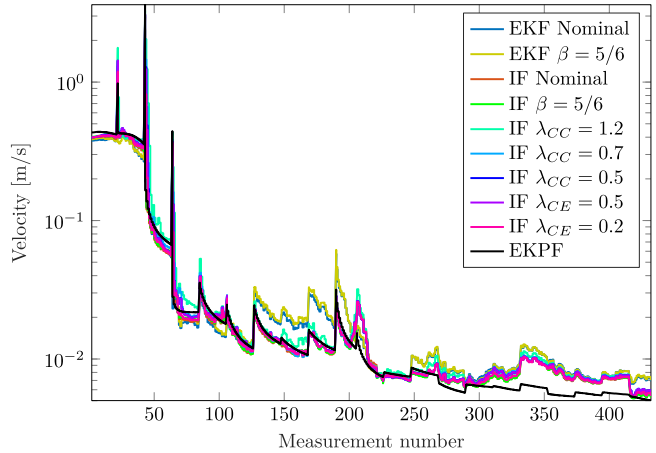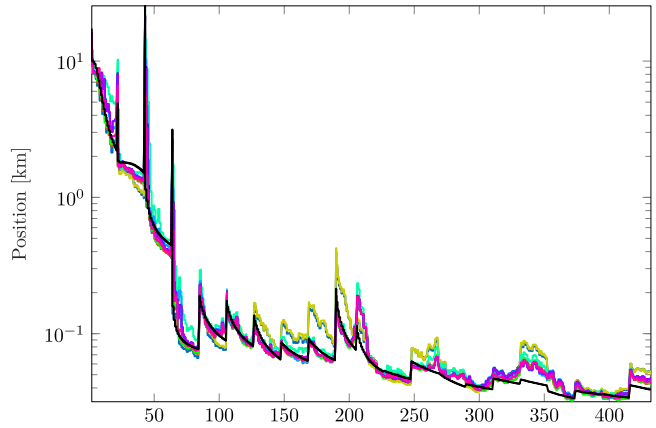
specific measurement number. The IFF with UW closely follows the nominal IFF with the exception of a handful of measurements over the course of the first two nights. At these exceptions, the UW conditions in (27) are met and UW is applied for that particular update resulting in very cautious filter performance until the next measurement where UW is not indicated, at which point the filter returns to consistent estimation performance.

The IFF simulations with either CC or CE rate-of-change show remarkably different behavior than that of the nominal or underweighted simulations. As with the traditional UW equivalent flow formulation, the convergence constrained filters show large drops in the number of filter estimate standard deviations required to fully capture the MC 3σ. While the UW simulation immediately returns to consistent estimation (i.e., values in the neighborhood of 3.0), the convergence control methods show a more gradual return to consistent estimation performance. The depth of the initial drop in the ratio value and the number of subsequent measurements processed before returning to consistent estimation scales directly with the λ_{CC} or λ_{CE} convergence threshold. That is, larger values of λ_{CC} or λ_{CE} result in more cautious filter performance and take more time to return to consistency than smaller values.

While some level of cautious performance is desirable, especially in situations of precise measurements coupled
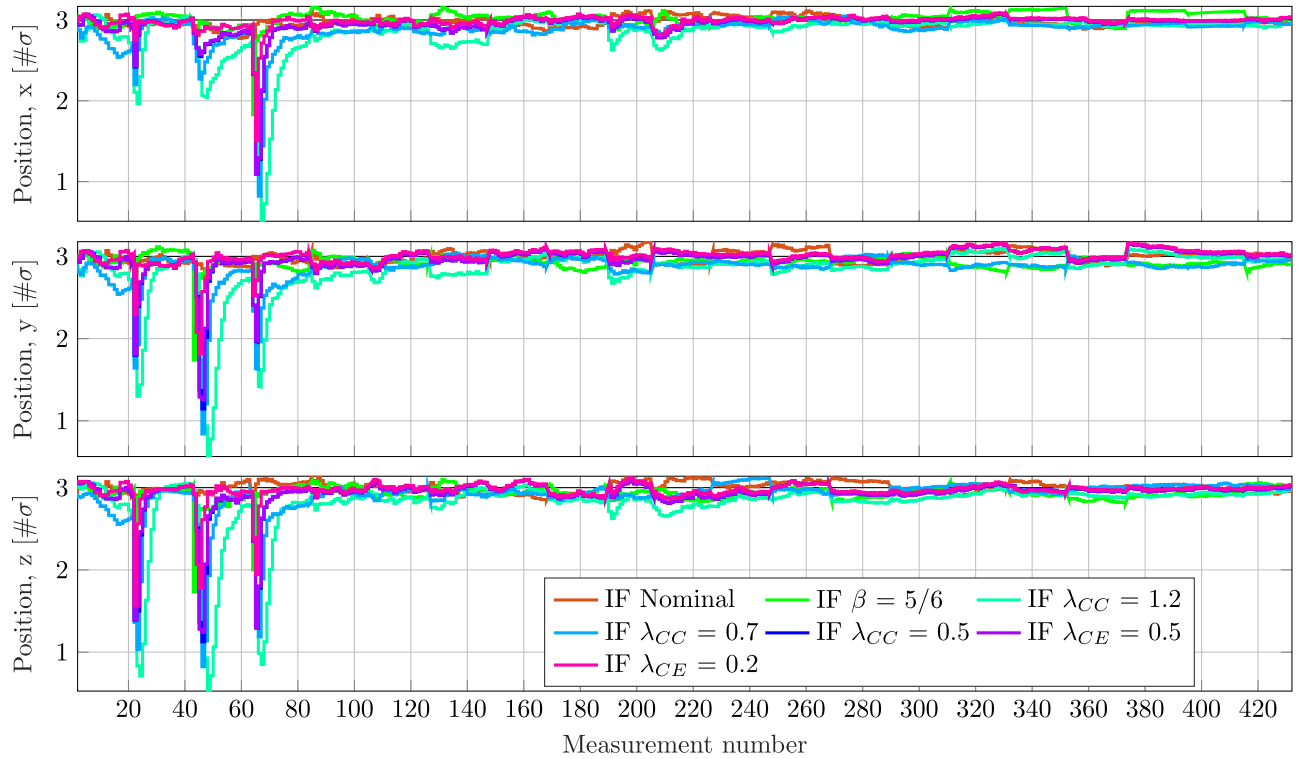
Fig. 8. Position results, average number of filter σ needed to capture MC error 3σ.

with high state uncertainty seen in this application, an over-abundance of caution can also be harmful. The tradeoff necessary to produce such caution in the filter response means that less information about the state is provided to the filter. As the amount of information withheld at a given update increases, so too does the number of subsequent updates required for the filter to return to consistent performance. To illustrate the consequences of this information deprivation, a consistency "score" is established for each simulation and estimated state. For the average ratio $\sigma_{\mathrm{avg},j}^{\ell}$ results [the simulation-specific lines in Figs. 8 and 9, computed using (33)], the sign-preserving squared offset from consistent performance in state j is summed for a simulation across the total $M = 432$ measurements and given by

$$S_j = \sum_{\ell=1}^{M} \mathrm{sign}(\sigma_{\mathrm{avg},j}^{\ell} - 3.0)(\sigma_{\mathrm{avg},j}^{\ell} - 3.0)^2 .$$

For simulations that are primarily cautious or smug in a given state, S_j will be negative or positive, respectively, with larger deviations from consistency further increasing the score magnitude. Ideally, filter scores should, therefore, be zero or slightly less than zero indicating, on average, consistent to slightly cautious estimation performance. Fig. 6 shows the scores for each IFF across the position and velocity states. The nominal IFF (Sim. 3) scores reflect its on-average consistent performance with small positive or negative scores depending on the specific state. Due to the amount of information withheld by the large convergence threshold, the IFF with $\lambda_{\mathrm{CC}} = 1.2$ (Sim. 5) shows scores

with consistently larger magnitudes than the other filter configurations across all states. The scores also reflect the trend of decreasing deviations from consistency as λ_{CC} or λ_{CE} decreases, with $\lambda_{\mathrm{CE}} = 0.2$ (Sim. 9) showing slightly cautious performance across all states. Note this does not mean that any given simulation was smug or cautious across all measurements, just in general. The IFF with UW (Sim. 4) varies in its scores relative to the convergence control-based filters depending on the state but does show cautious performance across all states. For the λ_{CC} and $\lambda_{\mathrm{CE}} = 0.5$ simulations (Sims. 7 and 8, respectively), the scores are nearly identical across all states with $\lambda_{\mathrm{CC}} = 0.5$ always of a slightly larger magnitude. This indicates the two methods provide similar consistency results for the same pseudotime convergence threshold despite the different formulations for computing the contraction efficiency.

The tradeoff for these more cautious or consistent filter designs comes at the cost of a slowed convergence rate. Fig. 7 demonstrates this penalty in terms of the average filter error root-sum-square (RSS) for the position and velocity states. During the first night of measurements (measurement number ≤ 63), the two EKF filters show lower RSS values than the IFFs for the majority of the time interval. Of course by the third night of measurements (measurement number ≥ 127), this is no longer the case as the EKFs struggle to maintain the solution after long periods of propagation. Comparison of the RSS trends across the IFFs reveals the cost of the more cautious behavior seen in Figs. 8 and 9, slower convergence toward consistent estimation directly corresponds to a slower rate of reduction in RSS values.
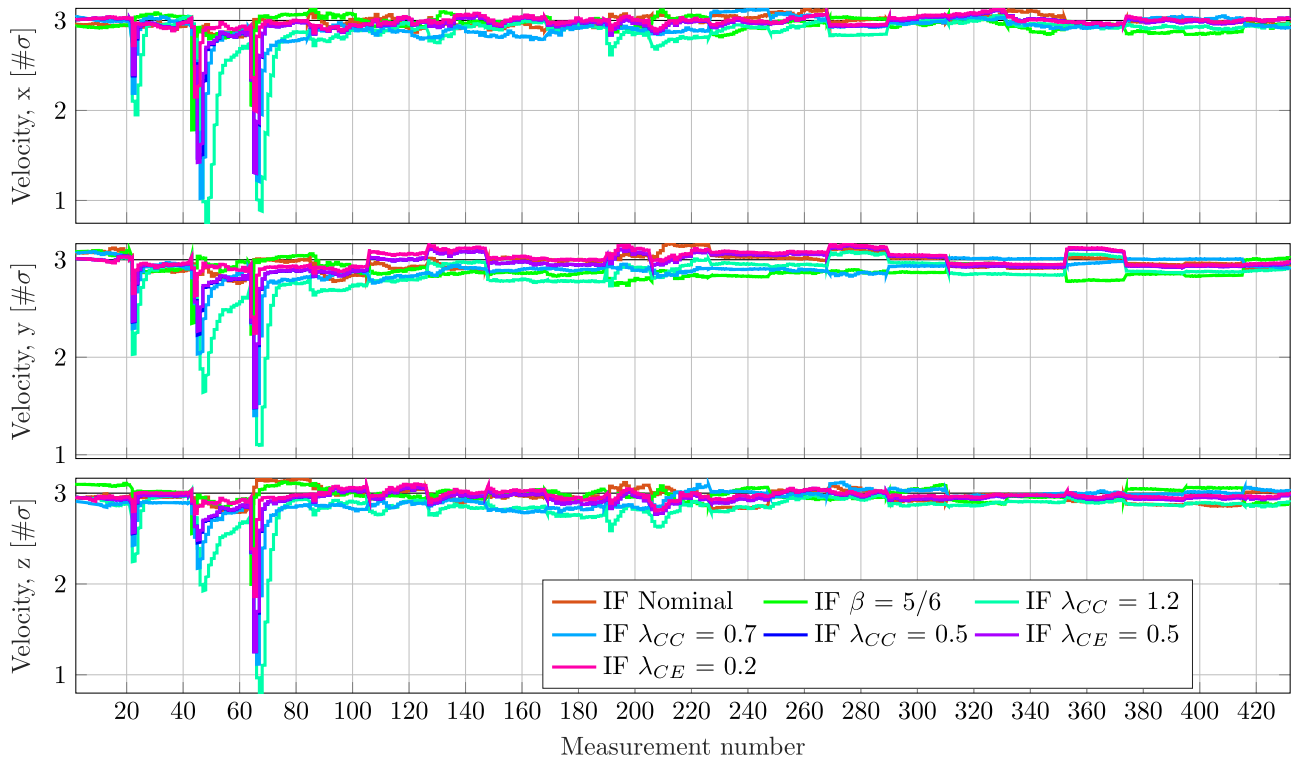
Fig. 9. Velocity results, average number of filter σ needed to capture MC error 3σ.

While all IF filters eventually steady state to the same RSS values, the number of measurements needed to achieve that steady state is determined by how much information is withheld from the filter when each measurement is processed. That is, the RSS values for each IFF at a given measurement decrease with the magnitude of the contraction efficiency, followed by the RSS values for the IF with $\beta = 5/6$ and then the nominal IF. RSS values for the GEPFF and GLEPFF are not shown as they are visually indistinguishable from the nominal IFF at this scale and a more detailed picture of the differences is already given in Fig. 3.

The EKPF RSS values are also given in Fig. 7. By and large, through the first eight nights (measurement number ≤ 268), the EKPF RSS values follow the same trends of those for the IFFs. Starting around measurement number 269, the EKPF demonstrates lower and lower RSS values for both position and velocity, compared to the other filters. Final position RSS values are 50 m and 49 m for the EKFs with and without UW, 43 m average for the IFFs ($\sigma = 1.1$ m), and 38 m for the EKPF. Final velocity RSS values are 6.9 mm/s and 7.0 mm/s for the EKFs with and without UW, 5.5 mm/s average for the IFFs ($\sigma = 0.16$ mm/s), and 5.0 mm/s for the EKPF.

The approximately 10% reduction in the position and velocity final RSS values of the EKPF compared to the average of the IFFs comes at the cost of a drastically higher computational burden. Using an i9-9900 K 3.60 GHz CPU with 32 GB RAM, the average MATLAB run time of the IFFs is 22.8 s for a single trial ($\sigma = 0.133$ s) with no distinguishing trends between the seven configurations. That

is, the presented convergence control methods do not add a significant computational burden compared to the nominal. The GEPFF, which relinearizes about the current particle set mean at each λ, results in a slightly higher average run time of 25.74 s ($\sigma = 0.2162$). The cost of frequent relinearization is even more pronounced in the GLEPFF with an average run time of 350.8634 s ($\sigma = 38.2790$). Conversely, the average run time for an EKF single trial is 1.70 s ($\sigma = 0.023$ s), resulting in a 13.4-fold increase in computation time from the EKF to the IFF. The single-trial EKPF run time, leveraging the MATLAB Parallel Computing Toolbox where appropriate for propagation and determining the unnormalized posterior weights, is 5 days 16.5 h. With the given computing conditions, this means that after nights with a higher number of measurements (e.g., nights 1, 2, 8, and 9), the EKPF would not finish processing the previous night's data before the next observations are received.

VI. CONCLUSION

A new method for robust and accurate estimation performance using an information-theoretic interpretation of the PFF has been presented. By motivating the movement of particles through the state space in accordance with a Bayesian update, particle flow methods have been shown to provide estimation accuracy similar to traditional particle filtering approaches at a reduced computational complexity. Deriving the flow using information theoretic measures allows for separation of the different modes that govern the motion of a particle through the update. Having distinct

parameters related to the modes of motion also allows for convenient implementation of a non-Bayesian update without significantly modifying the flow equations.

A novel pseudotime dependent contraction efficiency parameter is introduced by leveraging the distinction between the particle set contraction and translation over the course of an update. Based on either covariance or entropy rate-of-change, this parameter allows the filter to adapt the amount of information ingested at a given point in the update. Thresholding, and therefore slowing the rate of information inclusion into the estimate, prevents overconvergent behavior that can occur in situations with high state uncertainty and precise measurements.

The high uncertainty capabilities of the IFF with convergence control are demonstrated and compared to the EKF with and without UW, two Gaussian particle flow approaches, and the EKPF. The behavior is demonstrated under conditions that frequently overwhelm the EKF when the large state uncertainties encompass captured and non-captured orbits. The inclusion of a modifiable non-Bayesian update results in more desirable filter performance even with large initial uncertainties and gaps in time between measurements. Results indicate that the proposed IFF is capable of handling uncertainties that prove unmanageable for the EKF using a non-Bayesian update that is not readily implemented in other Gaussian particle flows. While the IFF has slightly higher error RSS values through the latter portion of the simulation compared to the EKPF, it is able to produce similar results with tunable levels of robust estimation in seconds, rather than days.

REFERENCES

[1] M. S. Grewal and A. P. Andrews
Kalman Filtering: Theory and Practice With MATLAB. Hoboken, NJ, USA: Wiley, 2014, ch. 6.

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

[3] N. Gordon, D. Salmond, and A. Smith
Novel approach to nonlinear/non-Gaussian Bayesian state estimation
IEE Proc. F - Radar Signal Process., vol. 140, no. 2, pp. 107–113, Apr. 1993.

[4] S. Särkkä
Bayesian Filtering and Smoothing. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[5] D. L. Alspach
A Gaussian sum approach to the multi-target identification-tracking problem
Automatica, vol. 11, no. 3, pp. 285–296, 1975.

[6] S. Särkkä
Rao-Blackwellized particle filter for multiple target tracking
Inf. Fusion, vol. 8, no. 1, pp. 2–15, 2007.

[7] K. J. DeMars and M. K. Jah
Probabilistic initial orbit determination using Gaussian mixture models
J. Guid., Control, Dyn., vol. 36, no. 5, pp. 1324–1335, 2013.

[8] F. Daum and J. Huang
Particle degeneracy: Root cause and solution
In *Proc. SPIE*, vol. 8050, May 2011, Art. no. 80500W.

[9] P. Bui Quang, C. Musso, and F. Le Gland
An insight into the issue of dimensionality in particle filtering
In *Proc. 13th Int. Conf. Inf. Fusion*, pp. 1–9, 2010.

[10] F. Daum and J. Huang
Nonlinear filters with particle flow induced by log-homotopy
Proc. SPIE, vol. 7336, pp. 76–87, 2009

[11] F. Daum, J. Huang, and A. Noushin
Exact particle flow for nonlinear filters
In *Proc. SPIE*, vol. 7697, pp. 92–110, 2010.

[12] A. Zellner
Optimal information processing and Bayes's theorem
In *Amer. Statistician*, vol. 42, no. 4, pp. 278–280, 1988.

[13] D. Guo, S. Shamai, and S. Verdu
Mutual information and minimum mean-square error in Gaussian channels
IEEE Trans. Inf. Theory, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[14] A. Wibisono, V. Jog, and P. Loh
Information and estimation in Fokker-Planck channels
In *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 2673–2677.

[15] J. E. Gough and N. H. Amini
Entropy production and information flow for Markov diffusions with filtering
2017, *arXiv:1710.05553*.

[16] F. Daum, J. Huang, and A. Noushin
Generalized Gromov method for stochastic particle flow filters
In *Proc. SPIE*, vol. 10200, Apr. 2017, Art. no. 102000I.

[17] C. Kreucher and K. Bell
A geodesic flow particle filter for non-thresholded measurements
In *Proc. IEEE Radar Conf.*, 2017, pp. 891–896.

[18] T. Ding and M. J. Coates
Implementation of the Daum-Huang exact-flow particle filter
In *Proc. IEEE Statist. Signal Process. Workshop*, 2012, pp. 257–260.

[19] F. De Melo, S. Maskell, M. Fasiolo, and F. Daum
Stochastic particle flow for nonlinear high-dimensional filtering problems
Mar. 2017, *arXiv:1511.01448*.

[20] F. Daum and J. Huang
Exact particle flow for nonlinear filters: Seventeen dubious solutions to a first order linear underdetermined PDE
In *Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, 2010, pp. 64–71.

[21] K. C. Ward and K. J. DeMars
Information-based particle flow for high uncertainty estimation
In *Proc. AIAA SciTech 2020 Forum*, 2020.

[22] C. E. Shannon
A mathematical theory of communication
Bell Syst. Tech. J., vol. 27, pp. 379–423, Jul. 1948.

[23] R. Zanetti, K. J. DeMars, and R. H. Bishop
Underweighting nonlinear measurements
J. Guid., Control, Dyn., vol. 33, no. 5, pp. 1670–1675, 2010.

[24] B. A. Kriegsman and Y.-C. Tao
Shuttle navigation system for entry and landing mission phases
vol. 12, pp. 213–219, no. 4, Apr. 1975.

[25] G. N. Holt, R. Zanetti, and C. N. D'Souza
Tuning and robustness analysis for the orion absolute navigation system
In *Proc. AIAA Guid., Navigat., Control Conf.*, 2013.

[26] K. J. DeMars, R. H. Bishop, and M. K. Jah
Entropy-based approach for uncertainty propagation of nonlinear dynamical systems
J. Guid., Control, Dyn., vol. 36, no. 4, pp. 1047–1057, 2013.

[27] F. Daum and J. Huang
A plethora of open problems in particle flow research for nonlinear filters, Bayesian decisions, Bayesian learning, and transport
In *Proc.* SPIE, vol. 9842, pp. 161–173, 2016.

[28] W. E. Wiesel
      Spaceflight Dynamics. New York, NY, USA: McGraw-Hill,
      1989.
[29] F. Daum and J. Huang
      Seven dubious methods to mitigate stiffness in particle flow with
      non-zero diffusion for nonlinear filters, Bayesian decisions, and
      transport
      In *Proc. SPIE*, vol. 9092, pp. 72–82, 2014.
[30] R. van der Merwe
      Sigma-point Kalman filters for probabilistic inference in dy-
      namic state-space models
      Ph.D. dissertation, OGI School Sci. & Eng., Oregon Health and
      Science University, Portland, Oregon, 2004.
[31] F. Daum and J. Huang
      The curse of dimensionality for particle filters
      In *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, Mar. 2003,
      pp. 4_1979–4_1993.

Kyle J. DeMars (Member, IEEE) received the
Ph.D., M.S.E., and the B.S. degrees in aerospace
engineering from The University of Texas at
Austin, Austin, TX, USA, in 2010, 2007, and
2004, respectively.

He is currently an Associate Professor of
Aerospace Engineering with Texas A&M Uni-
versity, College Station, TX, USA. His research
interests include Bayesian and non-Bayesian
filtering, nonlinear uncertainty prediction, or-
bit determination, data association, multitarget
tracking, attitude dynamics and determination, and autonomous sensor
management.

Kari C. Ward received the B.S. and Ph.D.
degrees in aerospace engineering from the
Missouri University of Science and Technology,
Rolla, MO, USA, in 2016 and 2021, respectively.

She is currently a Senior Member of the Tech-
nical Staff with the Charles Stark Draper Labora-
tory, Houston, TX, USA. Her research interests
include information theory, Kalman filtering,
sensor modeling, and vision-based spacecraft
navigation.