AIAA Guidance, Navigation, and Control Conference and Exhibit
16 - 19 August 2004, Providence, Rhode Island

AIAA 2004-5337

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

Particle Filtering for Sequential Spacecraft Attitude Estimation

Yang Cheng* and John L. Crassidis†

University at Buffalo, State University of New York, Amherst, NY 14260-4400

A new spacecraft attitude estimation approach using particle filtering is derived. Based on sequential Monte Carlo simulation, the particle filter approximately represents the probability distribution of the state vector with random samples. The filter formulation is based on the star camera measurements using a gyro-based or attitude dynamics-based model for attitude propagation. Modified Rodrigues parameters are used for attitude parametrization when the sample mean and covariance of the attitude are computed. The ambiguity problem associated with the modified Rodrigues parameters in the mean and covariance computation is addressed as well. By using the uniform attitude probability distribution as the initial attitude distribution and using a gradually decreasing measurement variance in the computation of the importance weights, the particle filter based attitude estimator possesses global convergence properties. Simulation results indicate that the particular particle filter, known as bootstrap filter, with as many as 2000 particles is able to converge from arbitrary initial attitude error and initial gyro bias errors as large as 4500 degrees per hour per axis.

I. Introduction

SEQUENTIAL attitude estimation falls in the category of nonlinear filtering. The Extended Kalman Filter (EKF) is the most widely used nonlinear filtering method for attitude estimation so far.[1] However, it works well only in the linear regime where the linear approximation of the nonlinear dynamical system is valid. Recently, an Unscented Filter (UF) has been proposed for attitude estimation, which uses a set of deterministically chosen *sigma points* to more accurately map a probability distribution, and which is more robust than the EKF under large initial attitude-error conditions.[2] Other improved algorithms for attitude estimation include the Extended Quaternion Estimator[3] and the Two-step Optimal Estimator,[4] both involving a combination of the standard Kalman Filtering technique with quadratically-constrained quadratic programming methods. All these three algorithms are based on second or higher order approximations of nonlinear functions and estimate the mean and covariance of the state vector. Though the mean and covariance are the sufficient statistics of a Gaussian distribution, they are not sufficient to represent a general probability distribution. The "global optimum" achieved by the Two-step Optimal Estimator also relies on a special sensor measurement model, namely, the attitude-vector measurement model. When these filtering methods are applied to strongly nonlinear and non-Gaussian estimation problems, where the posterior distribution of the state vector may be multi-peaked, heavily-tailed, or skewed, desired performance characteristics may not be obtained.

Attitude representation is an important issue in spacecraft attitude estimation. A spacecraft's attitude can be represented by a variety of parameters, including constrained (redundant) parameters such as the attitude matrix and the unit quaternion, and unconstrained (minimal) parameters such as the Euler angles, the Rodrigues parameters and the modified Rodrigues parameters (MRPs).[5] Constrained attitude parameters are singularity-free, but lead to constrained parameter estimation problems that will have to be solved directly[3,4] or circumvented[1,2] in attitude estimators. The most common approach in the EKF and the UF for ensuring the unit norm constraint of the attitude quaternion involves using two attitude representations: one is constrained and the other is unconstrained. The basic idea is to compute an unconstrained estimate of a three-component representation while using the correctly normalized quaternion to provide a globally

*Postdoctoral Research Associate, Department of Mechanical and Aerospace Engineering, cheng3@buffalo.edu.
†Associate Professor, Department of Mechanical and Aerospace Engineering, johnc@eng.buffalo.edu, Associate Fellow AIAA.

Copyright © 2004 by the American Institute of Aeronautics and Astronautics, Inc. All rights reserved.

nonsingular attitude representation.[6] The problem with unconstrained parameters is that no unconstrained parameters can be globally nonsingular, which limits the use of such parameters in representing the global attitude. The singularities associated with the MRPs, however, can be avoided based on the fact that for any physical attitude there are always two distinct sets of MRPs which never approach infinity simultaneously. Thus, in the sense of singularity avoidance, the MRPs may be the most preferable minimal attitude parameters. During attitude propagation, transforming the MRPs to the alternative set before they go too large can restrict the parameters within a specified sphere, far from singularity. These MRPs switches may or may not cause discontinuities of the MRPs error covariance, depending on how the error is defined. Although the ambiguity of the MRPs is desirable in singularity avoidance, it may cause problems in the expectation computations of the MRPs, as will be shown in a later section. The generalized Rodrigues parameters[7] have similar properties and problems as the MRPs.

In this paper a new attitude estimation approach based on the Particle Filter[8,9] (PF) is presented. Like other approximate approaches to optimal filtering, the ultimate objective of the PF is to construct the posterior probability density function (pdf) of the state vector, or the pdf of the state vector conditional on all the available measurements. However, the approximation of the PF is vastly different from that of conventional nonlinear filters. The central idea of the PF approximation is to represent a continuous distribution of interest by a finite (but large) number of weighted random samples of the state vector, or particles. The PF does not assume the posterior distribution of the state vector to be a Gaussian distribution or any other distribution of known form. In principle, it can estimate probability distributions of arbitrary form and solve any nonlinear and/or non-Gaussian system.

To some extent the UF[2,10] may be thought of as a very small-scale PF, since it also uses a small number of weighted "particles," named sigma points. However, the UF is essentially different from the PF in at least two aspects. First, the sigma points are deterministically computed from the mean and covariance of the state vector and are nothing but an equivalent representation of the mean and covariance. The weights on the sigma points are constant and pre-determined. The particles of the PF are randomly sampled from an *importance function*. The *importance weight* associated with each particle are adaptively computed based on the ratio between the posterior pdf and the importance function (up to a constant). Second, in the UF only the mean and covariance or the first two moments of the underlying distribution are estimated or maintained, whereas the PF attempts to approximate the underlying distribution directly by weighted particles. Given the particles, higher moments of interest as well as the mean and covariance can be computed in a straightforward manner whenever desired. From these particles, it is also convenient to compute statistics such as the modes and the median, which may be desired in certain applications. Stated in another way, the PF provides a whole picture of the underlying distribution that the UF cannot provide. In addition the UF allows use of negative weights, which makes the mathematical interpretation of a weighted average difficult.

A practical drawback of the PF is that sampling in high-dimensional state spaces can be inefficient. In order to remedy this drawback one technique, known as Rao-Blackwellization, reduces the size of the state vector to be estimated using the PF.[11] After marginalizing out part of the state vector analytically (using the linear Kalman filter in most cases), the PF only needs to handle the remaining part of the state vector that has a smaller size. But it is not always possible to Rao-Blackwellize the state space, since a special structure of the system is required. Another technique is based on annealing[12] or progressive corrections.[13] A heuristic idea is to introduce the influence of a sharp likelihood function gradually in a filter cycle, which is based on the adaptive decomposition of a sharp likelihood function into the product of several likelihood functions.

The particular PF proposed for attitude estimation is based on the Bootstrap Filter (BF), derived by Gordon, Salmond and Smith.[8] Being the first operational PF, the BF is modular and easy to implement. The justification for the BF is based on asymptotic results.[8] Thus, it is usually difficult to prove any general result for a finite number of samples or to make any precise, provable statement on how many samples are required to give a satisfactory representation of the pdf.[8] We prefer to use as few particles as possible in the BF, because the computational cost of the BF is largely proportional to the number of particles. For a BF with a modest number of particles to work properly, the sampling efficiency has to be enhanced. In order to do this, we include in the proposed BF the scheme of particle roughening, which was originally suggested in Ref. 8. More importantly, we choose to use a gradually decreasing measurement variance to handle the severe efficiency problem owing to the very accurate star camera measurements.

The organization of the remainder of this paper proceeds as follows. First, recursive Bayesian filtering and the BF are reviewed. Then, a brief review of the star camera/gyros measurement models as well as the

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

attitude motion model is provided. Next, a BF is derived for attitude estimation. Finally, the performance of the proposed BF is tested using simulated spacecraft data.

II. Bootstrap Filtering

In this section the theoretical framework of recursive Bayesian filtering and the BF are reviewed. A general discrete-time state-space model consists of the system model and the measurement model. The system model relates the current state vector, $\mathbf{x}_k$, to the one-stage ahead state vector, $\mathbf{x}_{k+1}$, and the measurement model relates the state vector to the measurement vector, $\tilde{\mathbf{y}}_k$:

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

$$\tilde{\mathbf{y}}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k) \tag{2}$$

In the above equations the system and measurement functions are denoted by $\mathbf{f}_k$ and $\mathbf{h}_k$, respectively. The subscript k in $\mathbf{f}_k$ and $\mathbf{h}_k$ indicates that the functions themselves can be time-varying. The measurement sampling period is denoted by $T = t_{k+1} - t_k$ and is assumed to be constant for simplicity. The vector $\mathbf{u}_k$ is the deterministic input. The process noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are assumed to be zero-mean white noise sequences. The distributions of the mutually independent $\mathbf{x}_0$, $\mathbf{w}_k$, and $\mathbf{v}_k$, denoted by $p(\mathbf{x}_0)$, $p(\mathbf{w}_k)$ and $p(\mathbf{v}_k)$, respectively, are assumed to be known. No Gaussian assumptions are required.

The central problem of Bayesian filtering or optimal filtering is to construct the posterior filtering distribution $p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_{k+1})$, where $\tilde{\mathbf{Y}}_{k+1} = \{\tilde{\mathbf{y}}_j : j = 0, \cdots, k+1\}$ is the set of measurements up to and including t_{k+1}. In addition we assume $p(\mathbf{x}_0|\tilde{\mathbf{Y}}_0) = p(\mathbf{x}_0)$. Under the above assumptions, the evolution of $p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_{k+1})$ only depends on the knowledge of the prior $p(\mathbf{x}_0)$, the transition $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and the likelihood $p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1})$. In principle, based on $p(\mathbf{w}_k)$ and the dynamical model Eq. (1), the transition $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ can be computed; based on $p(\mathbf{v}_k)$ and the measurement model Eq. (2), the likelihood $p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1})$ can be computed. The posterior filtering pdf $p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_{k+1})$ satisfies the following formal recursion:[8]

$$p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_{k+1}) = \frac{p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_k)}{\int p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_k)\, d\mathbf{x}_{k+1}} \tag{3a}$$

$$p(\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_k) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k)p(\mathbf{x}_k|\tilde{\mathbf{Y}}_k)\, d\mathbf{x}_k \tag{3b}$$

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k) = \int \boldsymbol{\delta}(\mathbf{x}_{k+1} - \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k))p(\mathbf{w}_k)\, d\mathbf{w}_k \tag{3c}$$

$$p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1}) = \int \boldsymbol{\delta}(\tilde{\mathbf{y}}_{k+1} - \mathbf{h}_{k+1}(\mathbf{x}_{k+1}, \mathbf{v}_{k+1}))p(\mathbf{v}_{k+1})\, d\mathbf{v}_{k+1} \tag{3d}$$

The quantity $\boldsymbol{\delta}(\cdot)$ in the above equations is Dirac's delta function and the above multi-dimensional integrals are defined as

$$\int \mathbf{f}(\mathbf{x})d\mathbf{x} = \int \cdots \int \int \mathbf{f}(\mathbf{x})dx_1 dx_2 \cdots dx_n \tag{4}$$

with $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$.

Except for very few dynamical systems such as linear Gaussian ones, the above recursive relations involve integrals that are mathematically intractable. The approximate solution the PF offers is based on Monte Carlo methods, in which a probability distribution is represented by a set of random samples. Given N independent and identically distributed random samples $\mathbf{x}^{(i)}$ drawn from $p(\mathbf{x})$, $i = 1, \cdots, N$, the distribution can be approximated by $p(\mathbf{x}) \approx (1/N)\sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}^{(i)})$ and an arbitrary integral (or expectation) with respect to $p(\mathbf{x})$ can be approximated by

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})d\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N} \mathbf{f}(\mathbf{x}^{(i)}) \tag{5}$$

Perfect Monte Carlo sampling assumes the samples are drawn directly from the distribution $p(\mathbf{x})$, but in practice it is seldom possible to do so. The PF is based on a sampling technique known as importance sampling. Rather than draw samples from the target distribution $p(\mathbf{x})$ directly, importance sampling draws $\mathbf{x}^{(i)}, i = 1, \cdots, N$, from an importance function $q(\mathbf{x})$ (also a pdf). These samples are weighted by the

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

normalized importance weights, which simultaneously satisfy $w^{(i)} \propto p(\mathbf{x}^{(i)})/q(\mathbf{x}^{(i)})$ and $\sum_{i=1}^{N} w^{(i)} = 1$, in order to account for the discrepancy between the importance function $q(\mathbf{x})$ and the target distribution $p(\mathbf{x})$. The samples drawn from an importance function and their importance weights $\{\mathbf{x}^{(i)}, w^{(i)}\}$ altogether form the two essential components of importance sampling. The integral in Eq. (5) is then approximated by $\sum_{i=1}^{N} w^{(i)}\mathbf{f}(\mathbf{x}^{(i)})$. For accuracy purposes, it is desired that an adequate number of samples drawn according to $q(\mathbf{x})$ are in high probability regions of $p(\mathbf{x})$ and the normalized importance weights are evenly distributed.

The PF is closely related to sequential importance sampling, an importance sampling method for recursive filtering and smoothing. The recursive form of sequential importance sampling requires a propagation mechanism between $\{\mathbf{x}_{k+1}^{(i)}, w_{k+1}^{(i)}\}$ at time t_{k+1} and $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}$ at time t_k. This is made possible by assuming that the importance function for $\mathbf{X}_k$ has the form

$$q(\mathbf{X}_{k+1}|\tilde{\mathbf{Y}}_{k+1}) = q(\mathbf{X}_k|\tilde{\mathbf{Y}}_k)q(\mathbf{x}_{k+1}|\mathbf{X}_k, \tilde{\mathbf{Y}}_{k+1}) \tag{6}$$

with $\mathbf{X}_k = \{\mathbf{x}_j : j = 0, \cdots, k\}$, the set of state vectors up to and including t_k. Note that such an importance function does not modify the previous particle trajectories. In a generic sequential importance sampling algorithm, the new particles at time t_{k+1}, $\mathbf{x}_{k+1}^{(i)}$, are drawn from an importance function $q(\mathbf{x}_{k+1}|\mathbf{X}_k^{(i)}, \tilde{\mathbf{Y}}_{k+1})$ and the importance weights on $\mathbf{x}_{k+1}^{(i)}$ are evaluated using

$$w_{k+1}^{(i)} \propto w_k^{(i)} \frac{p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1}^{(i)})p(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)})}{q(\mathbf{x}_{k+1}^{(i)}|\mathbf{X}_k^{(i)}, \tilde{\mathbf{Y}}_{k+1})} \tag{7}$$

A well-known problem with the sequential importance sampling method is the inevitable degeneracy phenomenon, in which after a few iterations, all but one particle will have negligible weights.[9] Therefore, it is of little practical use. In the BF a selection step (named resampling) is inserted after the importance weight update so as to reduce degeneracy. The basic idea of resampling is to discard particles with small weights and multiply particles with large weights while maintaining the total particle number unchanged. Because the resampling scheme always reduces the diversity of the particles, a roughening step is also added to our BF in order to increase the number of the distinct particles.[8]

The importance function $q(\mathbf{x}_{k+1}|\mathbf{X}_k^{(i)}, \tilde{\mathbf{Y}}_{k+1})$ in the BF is chosen as nothing but the prior $p(\mathbf{x}_{k+1}|\mathbf{x}_k^{(i)})$, independent of the previous particle trajectories before t_k and the measurements. One of the advantages of such a choice is that the importance weight is reduced to $w_{k+1}^{(i)} \propto w_k^{(i)} p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1}^{(i)})$, which only depends on the likelihood because $q(\mathbf{x}_{k+1}^{(i)}|\mathbf{X}_k^{(i)}, \tilde{\mathbf{Y}}_{k+1})$ and $p(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)})$ in Eq. (7) cancel each other out. Another advantage is that we only need to know how to draw samples from the prior $p(\mathbf{x}_{k+1}|\mathbf{x}_k^{(i)})$; we do not need to know how to evaluate it, which may be rather difficult. The disadvantage of the choice is also obvious. Because the generation of particles at time t_{k+1} depends on particles at time t_k and the system dynamics but does not take into account the measurement at time t_{k+1}, when the overlap between the prior and the likelihood is small, many particles may be propagated to regions of small likelihood and assigned with negligible importance weights. These particles will contribute little to the approximation of the posterior distribution or expectations. The approximation will be in effect dominated by only a small portion of particles with large weights. In other words the BF can be inefficient. But a better importance function or more efficient PF usually involves many more computations. So tradeoffs between sampling efficiency and computational cost always have to be made in practice.

In the ensuing the procedure of the BF is reviewed. Four steps, namely, prediction, update, resampling and roughening, constitute a filter cycle: from $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}$ to $\{\mathbf{x}_{k+1}^{(i)}, w_{k+1}^{(i)}\}$, where $i = 1, \cdots, N$. Note that although roughening is not an essential component of the basic BF, it is important in increasing particle diversity.

A. Prediction

The particles at time t_k are propagated through the following equation with their importance weights unchanged:

$$\mathbf{x}_{k+1}^{(i)} = \mathbf{f}_k(\mathbf{x}_k^{(i)}, \mathbf{u}_k, \mathbf{w}_k^{(i)}) \tag{8}$$

where N samples $\mathbf{w}_k^{(i)}$ of the process noise are drawn according to $p(\mathbf{w}_k)$, denoted by $\mathbf{w}_k^{(i)} \sim p(\mathbf{w}_k), i = 1, \cdots, N$.

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

B. Update

The importance weight associated with each particle is updated based on the likelihood function:

$$w_{k+1}^{(i)} = w_k^{(i)} p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1}^{(i)}) \tag{9a}$$

$$w_{k+1}^{(i)} \leftarrow \frac{w_{k+1}^{(i)}}{\sum_{i=1}^{N} w_{k+1}^{(i)}} \tag{9b}$$

where $\leftarrow$ denotes *set to* and the likelihood function $p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1}^{(i)})$ depends on the particular problem at hand. Note that $\sum_{i=1}^{N} w_{k+1}^{(i)} = 1$ after the normalization done by Eq. (9b).

C. Resampling and Roughening

The above prediction and update steps implement a cycle of the sequential importance sampling algorithm. For importance functions of the form of Eq. (6), the variance associated with the importance weights in sequential importance sampling can only increase over time, or eventually all but one particle will have negligible weight. A common practice to solve this degeneracy problem is to introduce resampling. Since the resampling scheme discards particles and may greatly decrease the number of distinct particles, the roughening procedure is followed to increase particle diversity.

The resampling and roughening steps may be applied at every cycle, as in the original BF. But it is not necessary to do so. The two steps are used in order to guarantee the proper performance of the BF, but are not required for processing the filter. The main point of resampling is to prevent the effective sample size, N_{eff}, from being too small. The disadvantage of resampling and roughening is that they introduce additional Monte Carlo variations. Also, in cases of less observability, applying these steps too frequently may even eliminate "good particle trajectories" that will have large weights for a longer data span. If resampling is done at every cycle, then Eq. (9a) reduces to

$$w_{k+1}^{(i)} = p(\tilde{\mathbf{y}}_{k+1}|\mathbf{x}_{k+1}^{(i)}) \tag{10}$$

The effective sample size is approximated by[9]

$$N_{\text{eff}} \approx 1/\sum_{i=1}^{N} (w_{k+1}^{(i)})^2 \tag{11}$$

which is a measure of variation of the (normalized) importance weights. If only very few particles have significant weight while others are negligible (the sum is always 1), then $N_{\text{eff}} \approx 1$; if all the particles are nearly equally weighted, then $N_{\text{eff}} \approx N$. To the extent that very small N_{eff} indicates severe diversity loss, large N_{eff} is desired. However, in the BF large N_{eff} alone does not necessarily ensure vast diversity among particles, because it may correspond to the unfavorable case in which most of the particles are identical (due to previous resampling steps).

Resampling is implemented by drawing samples (with replacement) N times from $\{\mathbf{x}_{k+1}^{(i)}, w_{k+1}^{(i)}\}$ to obtain N equally weighted particles, $\{\mathbf{x}_{k+1}^{(i)}, 1/N\}$. The number of particles remains unchanged after resampling. The normalized importance weight $w_{k+1}^{(i)}$ may be interpreted as the probability of occurrence for each particle. Stated in other words, the probability of the particle $\mathbf{x}_{k+1}^{(i)}$ being chosen at a single sample is approximately $w_{k+1}^{(i)}$ and after N samples $\mathbf{x}_{k+1}^{(i)}$ will be multiplied approximately $(N \cdot w_{k+1}^{(i)})$ times. The resampling algorithm is a black-box algorithm that takes as input the normalized importance weights and particle indices, and outputs new indices. It has nothing to do with the particles' dimension, values and so on. Of many resampling schemes, the particular scheme we choose is known as residual resampling.[14] It is a deterministic/random combined scheme and consists of two steps:

1. Retain $k_i = [N \cdot w_{k+1}^{(i)}]$ copies of $\mathbf{x}_{k+1}^{(i)}$ for each i, where $[\cdot]$ denotes the integer part of the number, e.g., $[32.3] = 32$ and $[12.6] = 12$. Let $N_r = N - k_1 - \cdots - k_N$.

2. Obtain N_r independent and identically distributed samples drawn from $\{\mathbf{x}_{k+1}^{(i)}, N \cdot w_{k+1}^{(i)} - k_i\}$, where the quantity $(N \cdot w_{k+1}^{(i)} - k_i)$ is just the fractional part of $N \cdot w_{k+1}^{(i)}$.

Note that in residual resampling only N_r samplings are needed. The residual resampling outperforms the simple random sampling scheme in having small Monte Carlo variance and favorable computational time.[14]

After resampling, roughening is done by[8]

$$\mathbf{x}_{k+1}^{(i)} \leftarrow \mathbf{x}_{k+1}^{(i)} + \mathbf{c}_{k+1}^{(i)} \tag{12}$$

with $\mathbf{c}_{k+1}^{(i)}$ an independent jitter drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0}, J_{k+1})$. The diagonal matrix J_{k+1} is denoted by $J_{k+1} = \mathrm{diag}([\sigma_1^2, \cdots, \sigma_n^2])$. The lth standard deviation σ_l is given by $\sigma_l = GE_l N^{-1/n}$, where E_l is the length of the interval between the maximum and the minimum samples of this component (before roughening), n is the dimension of the state space, and G is a tuning parameter. The correlation between components are not taken into account in this scheme. By taking the standard deviation of the jitter to be inversely proportional to the nth root of the sample size, the degree of roughening is normalized to the spacing between nodes of the corresponding uniform rectangular grid of N points.[8] The roughening step produces new particles and therefore increases particle diversity by additional artificial noise. The roughening parameter G cannot be too large or too small. Tradeoffs between spawning more distinct particles (large noise) and not altering the original distribution too much (small noise) has to be made.

In the BF, as well as other PFs, what is essential in the evolution of the filter is the particles and their associated importance weights. The derived practically significant quantities such as the mean and covariance only need to be computed at desired time points using the following equations:

$$\hat{\mathbf{x}}_{k+1} = E\{\mathbf{x}_{k+1}|\tilde{\mathbf{Y}}_{k+1}\} \approx \frac{1}{N}\sum_{i=1}^{N} w_{k+1}^{(i)}\mathbf{x}_{k+1}^{(i)} \tag{13a}$$

$$P_{k+1} = E\{(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1})^T|\tilde{\mathbf{Y}}_{k+1}\} \approx \frac{1}{N}\sum_{i=1}^{N} w_{k+1}^{(i)}\tilde{\mathbf{x}}_{k+1}^{(i)}\tilde{\mathbf{x}}_{k+1}^{(i)T} \tag{13b}$$

$$\tilde{\mathbf{x}}_{k+1}^{(i)} = \mathbf{x}_{k+1}^{(i)} - \hat{\mathbf{x}}_{k+1} \tag{13c}$$

There is no need to compute these quantities in order to process the filter. It is advised that when the mean and covariance are computed, they should be computed after the update but before resampling and roughening.[9] The reason is these two steps both introduce additional variations. The resampling step has an obvious "cut-tail" effect and the roughening step increases the sample covariance of the particles.

The BF makes few assumptions about the system and measurement models, and involves only straightforward function evaluations and random sampling schemes, thus being very easy to implement. The function evaluations include the system function, measurement function, and the likelihood function (possibly up to a constant). The following sampling steps are needed: drawing samples from $p(\mathbf{x}_0)$ at the initial time, drawing samples from $p(\mathbf{w}_k)$ at the prediction step, drawing samples from $\{\mathbf{x}_{k+1}^{(i)}, N \cdot w_{k+1}^{(i)} - k_i\}$ at the residual resampling step, and drawing samples from $\mathcal{N}(\mathbf{0}, J_{k+1})$ at the roughening step. Note that in the BF it is not required to draw samples from the likelihood or evaluate the prior.

Although it is discrete-time in nature, the BF is not limited to discrete-time/discrete-time state-space models. Its application can be extended to continuous-time/discrete-time state-space models with ease by converting the continuous-time model to its discrete-time counterpart. The prediction step is the only step involved in this conversion. This convenience is due to the fact that in the BF the prior and the likelihood are treated separately and that the evaluation of the prior is not required. Physical systems are usually described by a continuous-time model:

$$\frac{d\mathbf{x}(t)}{dt} = \boldsymbol{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t)) \tag{14}$$

where $\mathbf{w}(t)$ is a continuous-time zero-mean white noise process (it is also possible to assume that $\mathbf{w}(t)$ includes impulses). For linear Gaussian models, the equivalent discrete-time models are also linear Gaussian and the exact relationship between $\mathbf{x}_{k+1}$ and $\mathbf{x}_k$ can be solved for analytically. For general nonlinear non-Gaussian models, this relationship is usually obtained by numerically integrating the differential equations from t_k to t_{k+1}. If the right-hand side function contains no process noise, i.e., $\boldsymbol{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{0})$, the integration of Eq. (14) is straightforward. Otherwise, we need to determine how to approximate the white-noise process $\mathbf{w}(t)$ that does not exist in the real world. Here we use a band-limited, or piece-wise constant "white" noise $\mathbf{w}_h(t)$ as the approximation to $\mathbf{w}(t)$. The band-limited white noise $\mathbf{w}_h(t)$ changes its value only at time

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

$t_0, t_0+h, t_0+2h, \cdots$ and holds constant during the interval $[t_0+(k'-1)\cdot h, t_0+k'\cdot h)$. The small time interval is given by $h = T/K'$ with $K' \geq 1$ an integer. The sequence $\{\mathbf{w}_h(t_0+k'\cdot h)\}$, or $\{\mathbf{w}_{k'}\}$ for short, is a white-noise sequence with pdf $p(\mathbf{w}_{k'})$. Obviously, the pdf $p(\mathbf{w}_{k'})$ is related to the original $p(\mathbf{w}(t))$. For example, if $\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, Q)$, where Q is the spectral density, then we have $\mathbf{w}_{k'} \sim \mathcal{N}(\mathbf{0}, Q/h)$.

III. Attitude Motion and Sensor Models

The spacecraft attitude with respect to inertial space is parameterized by the MRPs, defined as $\boldsymbol{\rho} = \mathbf{n}\cdot\tan(\theta/4)$ (positive form) or $\boldsymbol{\rho}_s = -\mathbf{n}\cdot\cot(\theta/4)$ (negative form), where $\mathbf{n}$ and θ are the rotation axis and angle, respectively. The motivation to use MRPs in the BF based attitude estimator is discussed in detail in the next section. By definition the two sets of MRPs corresponding to the same physical attitude are related to each other by $\boldsymbol{\rho}_s = -\boldsymbol{\rho}/(\boldsymbol{\rho}^T\boldsymbol{\rho})$, or $\boldsymbol{\rho} = -\boldsymbol{\rho}_s/(\boldsymbol{\rho}_s^T\boldsymbol{\rho}_s)$ and $\|\boldsymbol{\rho}\|\cdot\|\boldsymbol{\rho}_s\| \equiv 1$, where $\|\boldsymbol{\rho}\| = \sqrt{\rho_1^2+\rho_2^2+\rho_3^2}$. For each physical attitude, if one set of MRPs satisfy $\|\boldsymbol{\rho}\| \geq 1$, then the other must satisfy $\|\boldsymbol{\rho}_s\| \leq 1$. The above relation shows that $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_s$ will never approach infinity simultaneously, which can be used for singularity avoidance. Another feature of the MRPs is that the two sets of MRPs are always distinct: $\|\boldsymbol{\rho}-\boldsymbol{\rho}_s\| \geq 2$. Both $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_s$ satisfy the same kinematical equations, i.e.,

$$\frac{d\boldsymbol{\rho}}{dt} = \frac{1}{4}B(\boldsymbol{\rho})\boldsymbol{\omega} \tag{15a}$$

$$\frac{d\boldsymbol{\rho}_s}{dt} = \frac{1}{4}B(\boldsymbol{\rho}_s)\boldsymbol{\omega} \tag{15b}$$

with

$$B(\boldsymbol{\rho}) = [(1-\boldsymbol{\rho}^T\boldsymbol{\rho})I_{3\times3} + 2[\boldsymbol{\rho}\times] + 2\boldsymbol{\rho}\boldsymbol{\rho}^T] \tag{16}$$

In these equations $\boldsymbol{\omega}$ is the absolute angular velocity, $I_{3\times3}$ is the identity matrix and $[\boldsymbol{\rho}\times]$ is the cross product matrix associated with $\boldsymbol{\rho}$, defined as

$$[\boldsymbol{\rho}\times] = \begin{bmatrix} 0 & -\rho_3 & \rho_2 \\ \rho_3 & 0 & -\rho_1 \\ -\rho_2 & \rho_1 & 0 \end{bmatrix} \tag{17}$$

In Eq. (15) either set of MRPs may become infinite as time increases. Therefore neither set of MRPs alone can be used to describe global attitude motion. But because the two sets will not approach the singular condition simultaneously, this singularity with the MRPs kinematical equations can be solved by using a switch from one set to the other, e.g., when $\|\boldsymbol{\rho}\| > 1$.

From Eq. (15) the mapping from $\boldsymbol{\rho}_{k'}$ at time $t_{k'}$ to $\boldsymbol{\rho}_{k'+1}$ at time $t_{k'+1}$ is approximately given by

$$\boldsymbol{\rho}_{k'+1} = \frac{(1-\boldsymbol{\rho}_{k'}^T\boldsymbol{\rho}_{k'})\tan\frac{\|\boldsymbol{\Theta}_{k'}\|}{4}\hat{\boldsymbol{\Theta}}_{k'} + (1-\tan^2\frac{\|\boldsymbol{\Theta}_{k'}\|}{4})\boldsymbol{\rho}_{k'} + 2\tan\frac{\|\boldsymbol{\Theta}_{k'}\|}{4}[\boldsymbol{\rho}_{k'}\times]\hat{\boldsymbol{\Theta}}_{k'}}{1+\tan^2\frac{\|\boldsymbol{\Theta}_{k'}\|}{4}\boldsymbol{\rho}_{k'}^T\boldsymbol{\rho}_{k'} - 2\tan\frac{\|\boldsymbol{\Theta}_{k'}\|}{4}\boldsymbol{\rho}_{k'}^T\hat{\boldsymbol{\Theta}}_{k'}} \tag{18}$$

with

$$\boldsymbol{\Theta}_{k'} = \int_{t_{k'}}^{t_{k'+1}} \boldsymbol{\omega}(\tau)d\tau \tag{19a}$$

$$\hat{\boldsymbol{\Theta}}_{k'} = \frac{\boldsymbol{\Theta}_{k'}}{\|\boldsymbol{\Theta}_{k'}\|} \tag{19b}$$

We assume that $\hat{\boldsymbol{\Theta}}_{k'} = \mathbf{0}$ when $\|\boldsymbol{\Theta}_{k'}\| = 0$. Note that the caret symbol in $\hat{\boldsymbol{\Theta}}_{k'}$ denotes a unit vector, not an estimate. Equation (18) is exact if during $[t_{k'}, t_{k'+1}]$ the direction of $\boldsymbol{\omega}(t)$ is fixed in the spacecraft body frame. When $\boldsymbol{\rho}_{k'}$ is parallel to $\hat{\boldsymbol{\Theta}}_{k'}$ and $\|\boldsymbol{\rho}_{k'}\|\cdot\tan\frac{\|\boldsymbol{\Theta}_{k'}\|}{4} = 1$, Eq. (18) is singular. Similarly, this singularity can be avoided by switching $\boldsymbol{\rho}_{k'}$ to the alternative set. An equivalent but singularity-free way to obtain $\boldsymbol{\rho}_{k'+1}$ is to propagate the attitude using the unit quaternion kinematics, and then convert $\mathbf{q}_{k'+1}$ to $\boldsymbol{\rho}_{k'+1}$ when needed, using the following equations:

$$\mathbf{q}_{k'+1} = \begin{bmatrix} \cos\frac{\|\boldsymbol{\Theta}_{k'}\|}{2}I_{3\times3} - [\hat{\boldsymbol{\Theta}}_{k'}\times]\sin\frac{\|\boldsymbol{\Theta}_{k'}\|}{2} & \hat{\boldsymbol{\Theta}}_{k'}\sin\frac{\|\boldsymbol{\Theta}_{k'}\|}{2} \\ -\hat{\boldsymbol{\Theta}}_{k'}^T\sin\frac{\|\boldsymbol{\Theta}_{k'}\|}{2} & \cos\frac{\|\boldsymbol{\Theta}_{k'}\|}{2} \end{bmatrix}\mathbf{q}_{k'} \tag{20a}$$

$$\boldsymbol{\rho}_{k'+1} = \frac{\boldsymbol{\varrho}_{k'+1}}{1+q_{4,k'+1}} \tag{20b}$$

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

In the above equations it is assumed that $\mathbf{q}_{k'+1} = [\boldsymbol{\varrho}_{k'+1}^T, q_{4,k'+1}]^T$ and $q_{4,k'+1} \geq 0$. Non-negative $q_{4,k'+1}$ guarantees $\|\boldsymbol{\rho}_{k'+1}\| \leq 1$.

Under rigid body assumption, the spacecraft dynamical equations are given by[15]

$$\frac{d\mathbf{H}}{dt} = -[\boldsymbol{\omega}\times]\mathbf{H} + \mathbf{T} \tag{21}$$

where $\mathbf{H} = \mathrm{I}\boldsymbol{\omega}$ is the spacecraft angular momentum and I is the inertia matrix. The vector $\mathbf{T}$ denotes all the torques exerting on the spacecraft and can be written as $\mathbf{T} = \mathbf{T}_0 + \mathbf{T}_{\text{dist}}$, where $\mathbf{T}_0$ and $\mathbf{T}_{\text{dist}}$ denote nominal and disturbance torques, respectively.

It is difficult for the attitude dynamical equations to provide a high precision attitude rate reference. A common sensor suite that measures the angular rate is rate-integrating gyros. For this sensor, a widely used model is given by[16]

$$\tilde{\boldsymbol{\omega}}(t) = \boldsymbol{\omega}(t) + \boldsymbol{\beta}(t) + \boldsymbol{\eta}_v(t) \tag{22a}$$

$$\dot{\boldsymbol{\beta}}(t) = \boldsymbol{\eta}_u(t) \tag{22b}$$

where $\tilde{\boldsymbol{\omega}}(t)$ is the continuous-time measured angular rate, and $\boldsymbol{\eta}_v(t)$ and $\boldsymbol{\eta}_u(t)$ are independent zero-mean Gaussian white-noise processes with

$$E\left\{\boldsymbol{\eta}_v(t)\boldsymbol{\eta}_v^T(\tau)\right\} = I_{3\times3}\sigma_v^2\delta(t-\tau) \tag{23a}$$

$$E\left\{\boldsymbol{\eta}_u(t)\boldsymbol{\eta}_u^T(\tau)\right\} = I_{3\times3}\sigma_u^2\delta(t-\tau) \tag{23b}$$

Attitude measurements are assumed to be given by a star camera. For simplicity, the following assumptions are made for the star camera measurement model: the spacecraft body frame and the camera frame coincide; the boresight of the camera is along the Z-axis of the camera frame; the camera of the star camera is an ideal pin-hole camera; and the measurement noise with each star is independent, stationary, zero mean and Gaussian with standard deviation σ_{ST}. The focal-plane representation of the star centroid is chosen as the basic data type of the star camera measurement model. It is directly related to the physical measurement and measurement noise. The measurement equations for the jth observed star in the field-of-view (FOV) at t_k are given as follows:

$$\tilde{\mathbf{y}}_k^j = \mathbf{p}_k^j + \boldsymbol{\varepsilon}_k^j \tag{24a}$$

$$\mathbf{p}_k^j = \begin{bmatrix} b_{1k}^j/b_{3k}^j \\ b_{2k}^j/b_{3k}^j \end{bmatrix} \tag{24b}$$

$$\mathbf{b}_k^j = A(\boldsymbol{\rho}_k)\mathbf{r}^j \tag{24c}$$

$$\mathbf{r}^j = \begin{bmatrix} \cos\alpha^j\cos\delta^j \\ \sin\alpha^j\cos\delta^j \\ \sin\delta^j \end{bmatrix} \tag{24d}$$

$$A(\boldsymbol{\rho}_k) = I_{3\times3} + \frac{-4(1-\boldsymbol{\rho}_k^T\boldsymbol{\rho}_k)[\boldsymbol{\rho}_k\times] + 8[\boldsymbol{\rho}_k\times]^2}{(1+\boldsymbol{\rho}_k^T\boldsymbol{\rho}_k)^2} \tag{24e}$$

where $\tilde{\mathbf{y}}_k^j$ and $\mathbf{p}_k^j$ are the respective measured and true (normalized) focal-plane representations associated with the LOS vector. The representation of the LOS vector in the inertial reference frame is denoted by $\mathbf{r}^j$ and the representation in the spacecraft body reference frame $\mathbf{b}_k^j$. The LOS $\mathbf{r}^j$ is computed based on a star catalog and is assumed to be fixed in inertial space and noise-free, which is a good approximation in attitude estimation. The quantities b_{1k}^j, b_{2k}^j, and b_{3k}^j are the X-axis, Y-axis and Z-axis components of $\mathbf{b}_k^j$, respectively, and $A(\boldsymbol{\rho}_k)$ is the attitude matrix parameterized by $\boldsymbol{\rho}_k$. The measurement noise $\boldsymbol{\varepsilon}_k^j$ is normally distributed with the pdf

$$p(\boldsymbol{\varepsilon}_k^j) \propto \exp(-\frac{\boldsymbol{\varepsilon}_k^{jT}\boldsymbol{\varepsilon}_k^j}{2\sigma_{\text{ST}}^2}) \tag{25}$$

Another star camera measurement model is the well-known QUEST measurement model for LOS mea-

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

surements,[17] which is equivalent to the focal-plane representation model to low order:

$$\tilde{\mathbf{b}}_k^j = \mathbf{b}_k^j + \mathbf{v}_k^j = A(\boldsymbol{\rho}_k)\mathbf{r}^j + \mathbf{v}_k^j \tag{26a}$$

$$\tilde{\mathbf{b}}_k^j = \frac{1}{\sqrt{1+\tilde{y}_{1k}^{j2}+\tilde{y}_{2k}^{j2}}}\begin{bmatrix}\tilde{y}_{1k}^j\\ \tilde{y}_{2k}^j\\ 1\end{bmatrix} \tag{26b}$$

with $\mathbf{v}_k^j \sim \mathcal{N}(\mathbf{0}_{3\times1}, \sigma_{\mathrm{ST}}^2(I_{3\times3} - \tilde{\mathbf{b}}_k^j\tilde{\mathbf{b}}_k^{jT}))$. The noise polluted measurement vector $\tilde{\mathbf{b}}_k^j$ is used in the covariance matrix because the true value $\mathbf{b}_k^j = A(\boldsymbol{\rho}_k)\mathbf{r}^j$ is not available in practice. Note that the covariance is singular; however, we can still have the pdf for $\mathbf{v}_k^j$:

$$p(\mathbf{v}_k^j) \propto \exp\left[-\frac{\mathbf{v}_k^{jT}(I_{3\times3} - \tilde{\mathbf{b}}_k^j\tilde{\mathbf{b}}_k^{jT})\mathbf{v}_k^j}{2\sigma_{\mathrm{ST}}^2}\right] \tag{27}$$

Because the measurement noise is additive, the derivation of the likelihood is straightforward. Take the focal-plane model as an example. Substituting $(\tilde{\mathbf{y}}_k^j - \mathbf{p}_k^j)$ for $\boldsymbol{\varepsilon}_k^j$ in Eq. (25) leads to the likelihood of $\boldsymbol{\rho}_k$:

$$p(\tilde{\mathbf{y}}_k^j|\boldsymbol{\rho}_k) \propto \exp\left[-\frac{(\tilde{\mathbf{y}}_k^j - \mathbf{p}_k^j)^T(\tilde{\mathbf{y}}_k^j - \mathbf{p}_k^j)}{2\sigma_{\mathrm{ST}}^2}\right] \tag{28}$$

Under the assumption that the measurement noise is independent, the joint likelihood at time t_k is $p(\tilde{\mathbf{y}}_k|\boldsymbol{\rho}_k) = \prod_{j=1}^M p(\tilde{\mathbf{y}}_k^j|\boldsymbol{\rho}_k)$, where the total measurement vector takes the form of $\tilde{\mathbf{y}}_k = [\tilde{\mathbf{y}}_k^1, \cdots, \tilde{\mathbf{y}}_k^M]^T$ and M is the number of observed stars used in attitude estimation.

IV. Bootstrap Filtering Based Attitude Estimator

In this section a general BF based attitude estimator is proposed for sequential spacecraft attitude estimation. We will consider two sensor settings. One involves the star camera and gyros and the other involves the star camera only. Star camera/gyros is one of the most accurate sensor suites for spacecraft attitude estimation; gyroless attitude estimation is desired in small spacecraft applications or for the case of gyro failures. These two settings lead to differences in the state variables and system model of the BF. In the star camera/gyros case, the state variables are the attitude and the gyro bias, where the system model is based on the attitude kinematics and the attitude rate is obtained from gyro measurements. In the gyroless case, the state variables are the attitude and the attitude rate, while the system model is given by the attitude kinematics and the attitude dynamics model. (For sake of simplicity only, we assume the attitude dynamics can be adequately modeled using Eq. (21).) The process noise in the two system models are the gyro measurement noise and the disturbance torque, respectively. The measurement model for the two settings are the same, i.e., the star camera measurement model, given by Eq. (24) or (26). Thus, the main difference between the two sensor settings only lies in the prediction step. Therefore, we can treat the two sensor settings in a general framework. In the following discussion we will concentrate on the star camera/gyros case. Three issues are addressed: attitude representations, initial attitude distribution and gradually decreasing measurement variance. The procedure of the BF is then presented at the end of the section.

A. Attitude Representations

In the BF the distribution of the attitude is represented by random samples of it. If the BF only consists of the prediction, update, and resampling steps, then any attitude representation or their combination can be used as long as the constrained attitude parameters obey the constraints and the unconstrained parameters avoid singularity. This is because the resampling step is independent of attitude representations, and in the prediction and update steps the only operations on the attitude parameters are the propagation of the attitude kinematics and the computation of the attitude matrix. Besides, transforming one attitude representation to another does not alter the importance weight on each particle.

The unit quaternion is the most widely used attitude representation in the EKF or the UF. Because the unit quaternion is singularity-free, subject to only one constraint, linear in its kinematical equations, and

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

bilinear in the attitude matrix, it may be the numerically best attitude representation for the prediction and update steps of the BF. But the unit quaternion is problematic in the computation of the mean and covariance and in the roughening step, because the operations (average/addition) in Eq. (12) or (13) will violate the unit norm constraint upon the unit quaternion. An un-normalized quaternion is meaningless as far as attitude representation is considered. The brute-force re-normalization of the un-normalized quaternion resulting from Eq. (12) or (13) may yield considerable errors. In addition the ambiguity of the unit quaternion (any attitude can always be parameterized by a pair of quaternions, $\pm\mathbf{q}$) also causes problems in Eq. (12) or (13). The unconstrained representation, however, can be used directly in the mean/covariance computation and roughening without causing any problems. We therefore choose to use the MRPs in the these steps:

$$\hat{\boldsymbol{\rho}}_{k+1} = \frac{1}{N}\sum_{i=1}^{N} w_{k+1}^{(i)} \boldsymbol{\rho}_{k+1}^{(i)} \tag{29a}$$

$$P_{k+1} = \frac{1}{N}\sum_{i=1}^{N} w_{k+1}^{(i)} \tilde{\boldsymbol{\rho}}_{k+1}^{(i)} \tilde{\boldsymbol{\rho}}_{k+1}^{(i)T} \tag{29b}$$

$$\tilde{\boldsymbol{\rho}}_{k+1}^{(i)} = \boldsymbol{\rho}_{k+1}^{(i)} - \hat{\boldsymbol{\rho}}_{k+1} \tag{29c}$$

$$\boldsymbol{\rho}_{k+1}^{(i)} \leftarrow \boldsymbol{\rho}_{k+1}^{(i)} + \mathbf{c}_{k+1}^{(i)} \tag{29d}$$

In Eq. (29) the MRPs are regarded as *real* vectors. By definition, the MRPs $\boldsymbol{\rho}_{k+1}^{(i)}$ in Eq. (29) are with respect to the inertial frame. We note that there is no substantial advantage to use the relative MRPs with respect to a predicted body frame when applying these equations, then compute the expectations with respect to the predicted body frame and finally convert the results back to the inertial frame. When the filter converges to its steady-state values, all the samples of MRPs must condense at a very small region around the true attitude, the means computed with respect to different frames are equivalent to first order. Before the BF converges to its steady-state values from a large initial error, the MRPs are sparsely spread in a rather large region of the parameter space, neither the inertial frame or the body frame may be a preferable reference frame for accuracy purposes. Similarly, we see no obvious advantage to roughen the particles, i.e., adding small jitters in the predicted body frame.

Because the mapping from the MRPs to the physical attitude is $2:1$, this ambiguity of the MRPs in the mean and covariance computation needs further discussion. Ignoring the ambiguity problem may result yield undesired results in the roughening step as well as mean/covariance computation. Either $\boldsymbol{\rho}_k^{(i)}$ or alternatively $\boldsymbol{\rho}_{sk}^{(i)}$ in Eq. (29a) are both admissible, but they will yield different results. The resulting means and covariances are not equally good. In order to see this consider a special example in which there are only two equally weighted particles, both corresponding to a rotation of π about the X-axis. The MRPs for the attitude are $[1,0,0]^T$ or equivalently $[-1,0,0]^T$. We expect the mean attitude is $[1,0,0]^T$ or $[-1,0,0]^T$, too. If no constraint on the MRPs is imposed, Eq. (29a) may reduce to $([1,0,0]^T + [1,0,0]^T)/2$, $([-1,0,0]^T + [-1,0,0]^T)/2$ or $([1,0,0]^T + [-1,0,0]^T)/2$. Obviously, $([1,0,0]^T + [1,0,0]^T)/2$ or $([-1,0,0]^T + [-1,0,0]^T)/2$ gives a much more reasonable result than $([1,0,0]^T + [-1,0,0]^T)/2 = [0,0,0]^T$, which is misleading. Similarly, the average $([1.1,0,0]^T + [0.9,0,0]^T)/2$ gives a more reasonable result than $([-1/1.1,0,0]^T + [0.9,0,0]^T)/2$. (In this example, $[1.1,0,0]^T$ and $[-1/1.1,0,0]^T$ correspond to the same attitude, i.e., an X-axis rotation of about 190°; $[0.9,0,0]^T$ corresponds to an X-axis rotation of about 170°. The average attitude is expected to be a rotation of about 180°.) The examples exclude the possibility of eliminating the ambiguity by simply confining the MRPs to $\|\boldsymbol{\rho}^{(i)}\| \leq 1$. Our approach (CONDMRP) to the ambiguity problem, which is outlined as follows, is to condense the samples of the MRPs as much as possible before computing the mean and covariance:

1. Of all the $\boldsymbol{\rho}_k^{(i)}$ and $\boldsymbol{\rho}_{sk}^{(i)}$, find $\boldsymbol{\rho}_{k,\mathrm{MIN}}$ which has the smallest norm;
2. For any particle, if $\|\boldsymbol{\rho}_k^{(i)} - \boldsymbol{\rho}_{k,\mathrm{MIN}}\| > \|\boldsymbol{\rho}_{sk}^{(i)} - \boldsymbol{\rho}_{k,MIN}\|$, then set $\boldsymbol{\rho}_k^{(i)}$ to $\boldsymbol{\rho}_{sk}^{(i)}$;
3. Compute the mean and covariance from the resultant particles.

The main point is to condense the particles and avoid artificial outliers caused by the MRPs ambiguity.

We also slightly modify the roughening step for the MRPs. In the original BF algorithm, E_l is calculated as the length of the interval between the maximum and the minimum samples of the lth state component,

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

e.g., for the X-component of the MRPs, $E_1 = \max(\rho_1) - \min(\rho_1)$. The modified formula is $E_l = E_l^+ + E_l^-$, where $E_l^+ = \max(\rho_l^+) - \min(\rho_l^+)$ and $E_l^- = \max(\rho_l^-) - \min(\rho_l^-)$. In the equations ρ_l^+ and ρ_l^- denote positive and negative values, respectively. This works better than the original form when the BF is far from convergence and the MRPs samples have disjointed peaks.

B. Initial Attitude Distribution

In the BF the initial attitude distribution is used to draw attitude samples. When the initial attitude errors are not large, e.g., $5°(1\sigma)$ for each axis, and no further information regarding the correlation between the three axes or the distribution is available, we may choose the initial attitude distribution as a zero-mean Gaussian distribution (in the body frame) with covariance $\mathrm{diag}[(5°)^2, (5°)^2, (5°)^2]$. If no initial attitude information is available, or the initial errors are very large, we may still approximate the initial attitude distribution by an Gaussian distribution with a very large covariance, as the practice in the EKF or the UF. But the covariance is generally a useful concept only for linear systems, which macroscopically the attitude is not.[18] Actually, the attitude error is bounded and the maximum attitude error corresponds to a rotation of π. Thus, too large a covariance is not reasonable physically. The covariance matrix in the EKF or the UF is actually used for data weighting. A large covariance matrix in the EKF or the UF may be conveniently interpreted as very small weight or as a loose constraint on the initial guess. For attitude sampling purposes, a very large variance does not yield the desired distribution. When no *a priori* attitude information is available, it is most desired to draw attitude samples from a uniform attitude distribution, especially when the number of particles is limited. Other distributions may overemphasize some regions of the attitude parameter space. The uniform distribution for the unit quaternion is constant over the whole surface of the 4-dimensional unit sphere.[18] Thus, generating a uniform attitude quaternion is equivalent to drawing samples from a uniform distribution defined over the surface of the 4-dimensional unit sphere. The procedure for drawing samples from the uniform attitude distribution is not unique. Because the Euler angles involves simpler pdfs, we choose to first draw N samples of the 3-1-3 Euler angles, and then convert the 3-1-3 Euler angles to the unit quaternion or MRPs. (It is also possible to use asymmetric Euler angles such as the 3-1-2 Euler angles.) The 3-1-3 Euler angles (ϕ, θ, ψ) are defined by successively rotating ϕ about the Z-axis, rotating θ about the X-axis and rotating ψ about the Z-axis. The uniform pdf for the 3-1-3 Euler angles is simply[18]

$$p_{313}(\phi, \theta, \psi) = p(\phi)p(\theta)p(\psi) \tag{30}$$

with $p(\phi) = p(\psi) = 1/(2\pi)$ and $p(\theta) = (\sin\theta)/2$. In addition it is assumed that $0 \le \phi \le 2\pi$, $0 \le \theta \le \pi$, and $0 \le \psi \le 2\pi$. The scheme for the generation of N uniform 3-1-3 Euler angles is given by[18]

1. Draw N samples $\phi^{(i)}$ from an independent uniform distribution on $[0, 2\pi)$.
2. Draw N samples $\mu^{(i)}$ from an independent uniform distribution on $[0, 1]$ and compute $\theta^{(i)} = \arccos(1 - 2\mu^{(i)})$.
3. Draw N samples $\psi^{(i)}$ from an independent uniform distribution on $[0, 2\pi)$.

The transformations from the 3-1-3 Euler angles to the unit quaternion or the MRPs can be found in Ref. 5.

C. Importance Weights

In the update step we choose to update the importance weights star by star. For the focal-plane representation model, we have

$$w_{k+1}^{(i),j} = w_{k+1}^{(i),j-1} \exp\left[-\frac{(\tilde{\mathbf{y}}_{k+1}^j - \mathbf{p}_{k+1}^{(i),j})^T(\tilde{\mathbf{y}}_{k+1}^j - \mathbf{p}_{k+1}^{(i),j})}{2\sigma_{\mathrm{ST}}^2}\right] \tag{31a}$$

$$w_{k+1}^{(i),j} \leftarrow \frac{w_{k+1}^{(i),j}}{\sum_{i=1}^N w_{k+1}^{(i),j}} \tag{31b}$$

where j runs from 1 to M, $w_{k+1}^{(i),0} = w_k^{(i)}$, $w_{k+1}^{(i)} = w_{k+1}^{(i),M}$, and $\mathbf{p}_{k+1}^{(i),j}$ is computed from the attitude sample $\boldsymbol{\rho}_{k+1}^{(i)}$ based on Eq. (24). We may use the QUEST model for the importance weight computation as well. (If

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

M is very large, e.g., on the order of 10, additional resampling and roughening steps may be inserted before all the stars are processed in order to have the importance weights more evenly distributed.)

The measurement models we have assumed so far do not take into account the FOV constraint. From the fact that $\mathbf{b}_{k+1}^{(i),j}$ and $-\mathbf{b}_{k+1}^{(i),j}$ will yield the same $\mathbf{p}_{k+1}^{(i),j}$, we can infer that the particles with incorrect attitude close to $-A_{k+1}$ may be inappropriately assigned with very large likelihood or importance weights using Eq. (31). The consequence is that those particles with falsely large weights will have the priority to be multiplied. As long as the attitude is observable and the BF performs normally, they will eventually be discarded when a series of star measurements are processed. But this problem of false likelihood may severely affect the capability of the BF to track the true attitude, for many particles may have been used to track the wrong attitude. A similar problem exists with the QUEST model. For instance, because of the projection operator $(I_{3\times 3} - \mathbf{b}_{k+1}^{j}\mathbf{b}_{k+1}^{jT})$ in the pdf, the attitudes yielding $\mathbf{b}_{k+1}^{j}$ and $-\mathbf{b}_{k+1}^{j}$ are considered to be equally good. Thus, we include an additional checking step, which is based on the observation that the component of $\mathbf{b}_{k+1}^{j}$ along the Z-axis must be close to $+1$. If $b_{3,k+1}^{(i),j}$ is smaller than a certain threshold, we set the corresponding importance weight to zero or a considerably small positive number.

The performance of the BF heavily relies on how many particles are in or are able to move to the regions of high (posterior) probability. In the initialization phase when the attitude errors and rate errors are large, we can only expect that the particles are sparsely distributed in a vast region and that few particles can be in the neighborhood of the true attitude and attitude rate. Thus, most of the deviations $(\tilde{\mathbf{y}}_{k+1}^{j} - \mathbf{p}_{k+1}^{(i),j})$ in Eq. (31) are tremendously large compared with the star camera accuracy. As an exponential function of the square of the norm of the deviation, the importance weight may be so small that it is numerically identical to zero. For example, in Matlab, $\exp(-30^2) = 0$. Thus, if $\|\tilde{\mathbf{y}}_{k+1}^{j} - \mathbf{p}_{k+1}^{(i),j}\| \geq 30\sigma_{\mathrm{ST}}$, the weight on the ith particle will numerically be zero. Suppose for star cameras $\sigma_{\mathrm{ST}} = 20$ arcsec, then $30\sigma_{\mathrm{ST}}$ is just 10 arcmin. The initial attitude error, however, may be several tens of degrees or even 180°. If all the particles have numerically zero weight, we have no way to discriminate or update them. Even if this is not the case, there can only be very few particles that have considerable weight and that can survive after resampling. In other words, after resampling the particle population will quickly collapse to a few distinct particles, which are not necessarily close to the true states. The roughening step can only increase diversity in a very limited way, because the jitter covariance for roughening directly depends on the range of the values of the particles, which is very small in this case. Unable to keep an adequate number of distinct particles around the true states, the BF will eventually lose the capability to track the measurements or the true states. There is probably no easy way to solve the problem due to the vast difference between the high accuracy of the star camera and the very large attitude errors.

We propose to use a time-varying $\sigma_{\mathrm{ST}}(t)$ that gradually decreases from an adequately large value at time 0 to the final true value σ_{ST} at some t_f, e.g., from 30° to 20 arcsec. Large $\sigma_{\mathrm{ST}}(t)$ during the initialization phase allows many more particles a chance to pass the less stringent scrutiny of resampling and to be reproduced. Apparently, the measurement data before t_f are underweighted and therefore not fully utilized; this introduces extra estimation errors and leads to longer convergence time. However, the simpleness and effectiveness of the technique may outweigh its sub-optimality. Because of the large initial errors, the filter will not converge to the correct solution until an adequate number of measurements are processed to eliminate the ambiguity in the state estimates. From the perspective of the parameter space search, the data before t_f are used to direct the particles to the neighborhood of the true state vector so as to provide good *a priori* knowledge for later processing. Once the particles are close enough to the true states, the true standard deviation σ_{ST} can be safely used because the process noise in our problem is small. So, it will not drive the particles around the true states far away during a star camera sampling period. (If the process noise is very large or the sampling period is very long, and the direct use of σ_{ST} in the importance weight computation causes the inefficiency problem, we may need to consider using progressive corrections[13] during each cycle.)

D. Bootstrap Filter

The BF for star camera/gyros based attitude estimation is now summarized. The state vector is defined by $\mathbf{x} = [\boldsymbol{\rho}^T, \boldsymbol{\beta}^T]^T$, where $\boldsymbol{\rho}$ is the MRPs and $\boldsymbol{\beta}$ is the gyro bias. The procedure in the BF is as follows.

1. Initialization: At time 0, draw N samples $\boldsymbol{\rho}_0^{(i)}$ of the MRPs and N samples $\boldsymbol{\beta}_0^{(i)}$ of the gyro bias from the initial distribution. The initial attitude distribution is taken to be a Gaussian or a uniform attitude distribution. The uniform attitude distribution is used when the *a priori* attitude knowledge

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

is unavailable. The initial bias distribution is approximated by a Gaussian distribution, if no further knowledge is available.

2. From t_k to t_{k+1}, $k = 0, 1, 2, \cdots$

   Propagate $\boldsymbol{\rho}_k^{(i)}$ and $\boldsymbol{\beta}_k^{(i)}$ to obtain $\boldsymbol{\rho}_{k+1}^{(i)}$ and $\boldsymbol{\beta}_{k+1}^{(i)}$. The gyro measurement sampling period h is much smaller than the star tracker measurement sampling period T. We assume they satisfy $K' = T/h$, where K' is an integer. For each particle, draw K' samples $\boldsymbol{\eta}_{u,k'}^{(i)} \sim \mathcal{N}[\mathbf{0}_{3\times1}, \sigma_u^2 h I_{3\times3}]$, and K' samples $\boldsymbol{\eta}_{v,k'+1}^{(i)} \sim \mathcal{N}[\mathbf{0}_{3\times1}, \sigma_v^2/h I_{3\times3}]$, where $k' = 1, \cdots, K'$. The two sequences $\boldsymbol{\eta}_{u,k'}^{(i)}$ and $\boldsymbol{\eta}_{v,k'+1}^{(i)}$ are used for an approximation of the processes $\boldsymbol{\eta}_u(t)$ and $\boldsymbol{\eta}_v(t)$ in Eq. (22), with $t \in [t_k, t_{k+1}]$. The rate $\boldsymbol{\omega}_{k'+1}^{(i)}$ and the bias $\boldsymbol{\beta}_{k'+1}^{(i)}$, are propagated using

$$\boldsymbol{\omega}_{k'+1}^{(i)} = \tilde{\boldsymbol{\omega}}_{k'+1} - \boldsymbol{\beta}_{k'+1}^{(i)} - \boldsymbol{\eta}_{v,k'+1}^{(i)} \tag{32a}$$

$$\boldsymbol{\beta}_{k'+1}^{(i)} = \boldsymbol{\beta}_{k'}^{(i)} + \boldsymbol{\eta}_{u,k'}^{(i)} \tag{32b}$$

   Next, compute $\boldsymbol{\Theta}_{k'}^{(i)} = \boldsymbol{\omega}_{k'}^{(i)} h$ (or $\boldsymbol{\Theta}_{k'}^{(i)} = \frac{\boldsymbol{\omega}_{k'}^{(i)} + \boldsymbol{\omega}_{k'+1}^{(i)}}{2} h$), and $\hat{\boldsymbol{\Theta}}_{k'}^{(i)} = \frac{\boldsymbol{\Theta}_{k'}^{(i)}}{\|\boldsymbol{\Theta}_{k'}^{(i)}\|}$, then $\boldsymbol{\rho}_{k'+1}^{(i)}$ can be propagated from $\boldsymbol{\rho}_{k'}^{(i)}$ by applying Eq. (18) K' times. The CONDMRP operation is applied on $\boldsymbol{\rho}_{k'+1}^{(i)}$.

3. Compute the importance weight for each particle using Eq. (31). A time-varying $\sigma_{\mathrm{ST}}(t)$ is used for the case of large initial errors. If $b_{3,k+1}^{(i),j}$ is smaller than a certain threshold, we set the corresponding importance weight to zero or a considerably small positive number.

4. Compute the means $\hat{\boldsymbol{\rho}}_{k+1}$ and $\hat{\boldsymbol{\beta}}_{k+1}$ and the associated covariances using Eq. (13). The attitude rate estimate is given by $\hat{\boldsymbol{\omega}}_{k+1} = \tilde{\boldsymbol{\omega}}_{k+1} - \hat{\boldsymbol{\beta}}_{k+1}$.

5. If resampling and roughening is needed, use the residual resampling scheme to obtain N equally weighted particles. Roughen the equally weighted particles using Eq. (12). We choose to use the parameter R-factor to control how often the resampling and roughening steps are applied. For instance, R-factor $= 5$ means these steps are applied every 5 filter cycle. The parameter G in the roughening step is another free parameter.

Note that if no stars are available at time t_{k+1}, steps 3 and 5 are skipped. For the gyroless case, attitude rate propagation in step 2 is implemented by integrating Eq. (21). Similarly, the continuous-time process noise $\mathbf{T}_{\mathrm{dist}}$ in Eq. (21) is approximated by a discrete-time sequence $\mathbf{T}_{\mathrm{dist},k'}$.

The above is a pure MRPs attitude estimator, in which no other attitude representations are used. An alternative estimator is such that the unit quaternion is used in the initialization, prediction, update and resampling steps, and the MRPs are used only in the mean/covariance computation and roughening steps. The two estimators will always yield the same results. The latter may be more computational efficient.

V. Simulation Examples

In this section the accuracy and convergence properties of the proposed BF are tested and compared with those of the UF using numerical simulations. For better comparisons, the attitude error in the UF is defined as the algebraic difference between the true and estimated MRPs relative to the inertial frame, the same as in the BF. The state vector of the UF is augmented to incorporate the process noise so that the effect of the nonlinear-appearing noise can be directly used. The square root decomposition of the covariance matrix is implemented using an eigenvalue decomposition (or SVD decomposition). The small negative eigenvalues of the covariance matrix due to numerical errors are set to zero to guarantee positive semi-definiteness. Before the predicted mean and covariance are to be computed from the sigma points, these are checked and adjusted using the CONDMRP procedure.

The inertia matrix of the spacecraft is assumed to be $\mathrm{I} = \mathrm{diag}[60, 60, 40]\mathrm{kg} \cdot \mathrm{m}^2$. For simplicity, the nominal torque $\mathbf{T}_0$ is assumed to be zero and the external disturbance torque $\mathbf{T}_{\mathrm{dist}}$ is modeled as a Gaussian distribution with a standard deviation of 5×10^{-5} Nm for each axis. The star camera and gyros are sampled at 2 seconds and 0.4 seconds intervals, respectively. The star camera measurements are simulated based on the star catalog and properties of the camera. The FOV of the star camera is $20^\circ \times 20^\circ$, the sensitivity

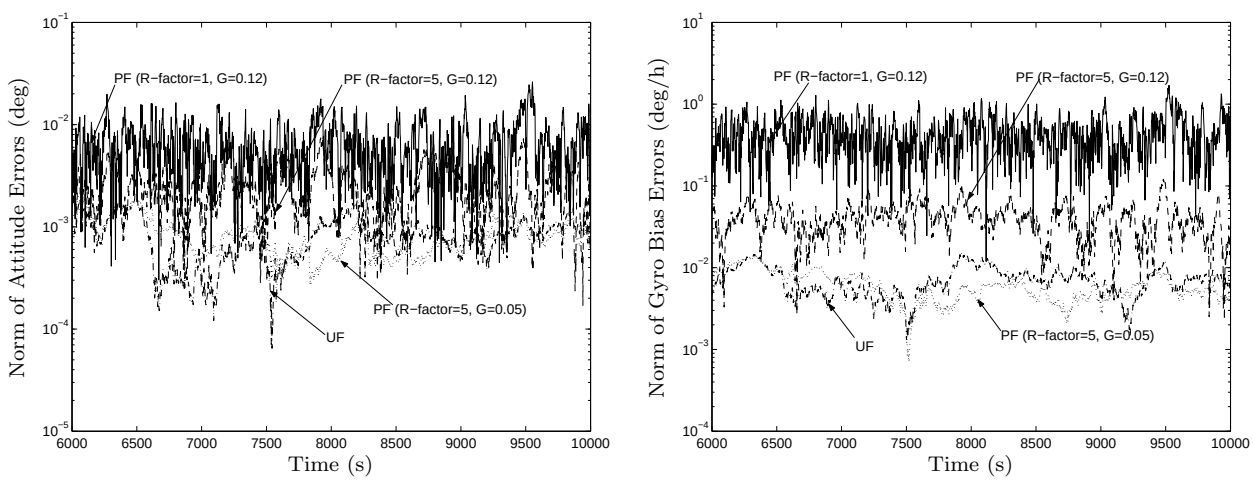
Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

threshold $M_{\text{threshold}} = 4.5$ Mv and the accuracy $\sigma_{\text{ST}} = 20$arcsec. The maximum number of stars used for the filter update is 5. If more than 5 stars are sensed, only the 5 brightest stars are used, regardless of the geometry (for sake of simplicity). The parameters for the gyros are $\sigma_u = 3.1623 \times 10^{-4} \mu\text{rad/s}^{3/2}$, and $\sigma_v = 0.31623 \mu\text{rad/s}^{1/2}$. The initial gyro bias is always modeled as a zero-mean Gaussian distribution with the standard deviation for each axis 0.1 deg/hr.

First the accuracy of the BF and the UF is examined. The tests are limited to the star camera/gyros setting only. The true values for the initial attitude and the initial attitude rate are $[0.2, 0.2, 0.2]^T$ (in MRPs) and $[0.002, -0.002, 0.002]^T$ rad/s, respectively. The initial distributions of the attitude and the gyro bias in the BF or the UF are taken to be Gaussian. The means are assumed to be the true values. The covariance matrices for both the initial attitude and the initial gyro bias are assumed to be isotopic, with the standard deviation 1 arcmin and 1 deg/hr, respectively. The data span is 10000 seconds or 5000 data points. Before the comparisons, we note that in this nearly linear error case the accuracy of the UF estimates is largely independent of the choice of its parameters, namely, α, β and κ (see Ref. 19 for a description of these variables). Thus, efforts on parameter tuning are not necessary. Because in these tests the initial errors and the process noise are small, the UF (or even the EKF) can perform in an almost optimal manner (strictly optimal for linear Gaussian systems), whereas the BF can never be strictly optimal, even if the system is linear and Gaussian. That is because based on Monte Carlo simulation, the BF always introduces extra errors in the state estimates. Thus, we should not expect the BF to yield better accuracy than the UF in these tests. In order to quantify the simulation results, the error norms are summed over the last 4000 seconds or 2000 data, given by

$$J = \sum_{k=3000}^{5000} \|\mathbf{e}_k\| \tag{33}$$

where $\mathbf{e}_k$ denotes the attitude errors or gyro bias errors. Finally, Eq. (33) is averaged over 12 runs. The steady-state errors of a typical run is given in Figure 1.



(a) Attitude Errors

(b) Gyro Bias Errors

Figure 1. Accuracy of the BF

The use of different star camera measurement models in either the UF or the BF does not lead to distinguishable results, with the difference over one order below the estimation accuracy. So the averaged J values are given without specifying which measurement model is used. The averaged J values for the UF are 1.35° and 12.0 deg/hr, independent of the UF parameters and approximately the theoretical lower bounds. In the BF, 2000 particles are used. When the resampling and roughening steps are applied every filter cycle (R-factor = 1), and G is set to 0.12, the averaged J values are 12.20° and 857.4 deg/hr. These values are about 9 times and 70 times their counterparts for the UF. However, the averaged J values reduce to 4.76° and 78.3 deg/hr when the particles are resampled and roughened every five cycles (R-factor = 5). When the resampling and roughening steps are applied every five filter cycles (R-factor = 5) and G is set to 0.05, the averaged J values for the BF become 1.45° and 13.15 deg/hr, less than 10% larger than the average J

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

values for the UF. Clearly, fewer resampling and roughening steps as well as smaller G can greatly reduce the estimation errors in the BF. But too few resampling and roughening steps and too small G may cause very poor performance or divergence of the BF, even though the initial estimation errors are small in these tests. In short, the accuracy of the BF is lower than that of the UF and is greatly influenced by the resampling and roughening steps. The parameters for these two steps, namely, R-factor and G, need to be tuned in order to achieve a balance between convergence and variance reduction. Finally we note that the accuracy of the BF may be increased by employing more particles, however, even in perfect Monte Carlo sampling, the estimation accuracy is only proportional to $1/\sqrt{N}$, thus this approach for accuracy enhancement is inefficient and is therefore not studied.

The real advantage of the BF is when large initial errors exist. Now the convergence properties of the BF for star camera/gyros based attitude estimation are to be studied. At most five stars are used for each update. The true value for the initial attitude in each run is randomly drawn from the uniform attitude distribution. In each run, the true value for the initial attitude rate is randomly drawn from a zero-mean Gaussian distribution with the standard deviation for each axis 0.03 rad/s. In the BF the uniform attitude distribution is taken as the initial attitude distribution. The initial gyro bias is modeled as a Gaussian distribution with both the mean and the standard deviation per axis given by 1500 deg/hr. This filter setting corresponds to the case of no *a priori* attitude knowledge and very poor gyro bias estimates. (The true value for the gyro bias per axis is within 0.3 deg/hr.) We still use 2000 particles in the BF. The parameter G for roughening is set to 0.12. The time-varying $\sigma_{\mathrm{ST}}(t)$, designed for the worst case (with the largest initial errors) using trial and error, is given in Table 1. The first row of the table gives the time points

Table 1. Time-Varying Measurement Standard Deviation

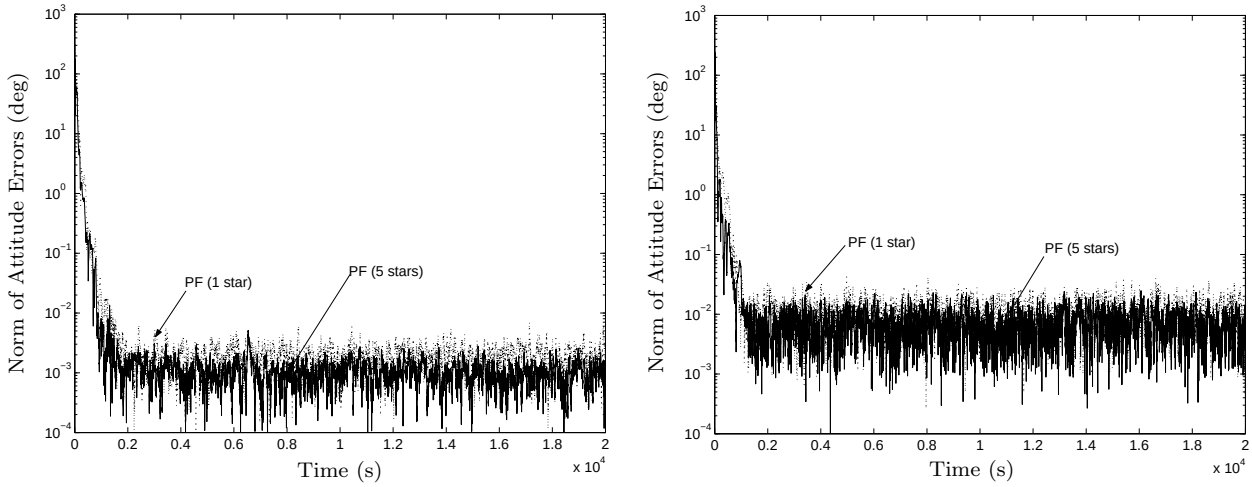
t	0	100s	200s	400s	500s	600s	700s	800s	1000s	1100s	1200s	1300s
$\sigma_{\mathrm{ST}}(t)$	20°	10°	4°	3°	2°	1°	0.5°	0.2°	0.1°	0.06°	0.02°	20″

at which the measurement standard deviation decreases its value and the second row gives the post-decrease standard deviation values. The value of the standard deviation $\sigma_{\mathrm{ST}}(t)$ decreases to 20arcsec in 650 star camera sampling periods, or 1300 seconds. There is no strict requirement on how often the resampling and roughening steps should be applied, but we find that initially (e.g., for the first 30 measurements) we have to do resampling and roughening every filter cycle (R-factor = 1) in order to suppress the particle degeneracy and diversity loss, which are most severe at the initial times. For instance, after the first update, the resulting normalized importance weights may be of large variance (i.e., only a few particles have considerable weights) and the effective number of particles may be only a few hundred or less. Without immediate resampling and roughening, the BF may fail to perform properly at later times. After the BF converges, we can keep using R-factor = 1 or reduce the frequency of resampling and roughening. The simulation for the star camera/gyros setting is run 100 times. In all runs the BF converges to the correct solutions. The result of the worst case in which the initial attitude estimate error is 180° and the initial gyro bias estimate error is $[4500, -4500, 4500]^T$ deg/hr is given in Figure 2.

We also study the special case in which a single star (the brightest star if there are stars in the FOV) is used for an update. In this single star case, we employ fast body rates (e.g., a spinning spacecraft) and force the use of different stars in different frames when possible. (For example, if another star comes into the FOV in a successive frame, then we use it instead of a previous star that may still be in the FOV.) Thus, the reference star "field" moves considerably with respect to the inertial frame. Therefore, observability of three-axis attitude is possible, akin to the magnetometer-only attitude estimation approach. The roughening parameter G and the decreasing $\sigma_{\mathrm{ST}}(t)$ in the single star case is the same as in the multi-star case. But it is observed that if R-factor = 1 is used during the whole simulation run, the BF may sometimes diverge. Therefore, except for setting R-factor = 1 for the first few seconds (e.g., 60 seconds), R-factor = 5 is used. Less resampling and roughening may be able to effectively prevent the BF from tracking the plausible local minima too much. (If most particles of the BF are used to track an incorrect attitude solution, the BF will quickly diverge.) Thus, the rule of thumb for tuning the R-factor is to initially set R-factor = 1 and afterwards a larger R-factor is preferred for better convergence properties. Another 100 runs are executed. Again, the BF converges to the correct solution in all the runs. The result for the worst case is also given in Figure 2. We also test the convergence properties of the BF for the gyroless setting. The attitude dynamical model needed for the attitude rate propagation in the gyroless setting is "more" nonlinear than the gyro-

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337



(a) Star Camera/Gyros

(b) Star Camera Alone

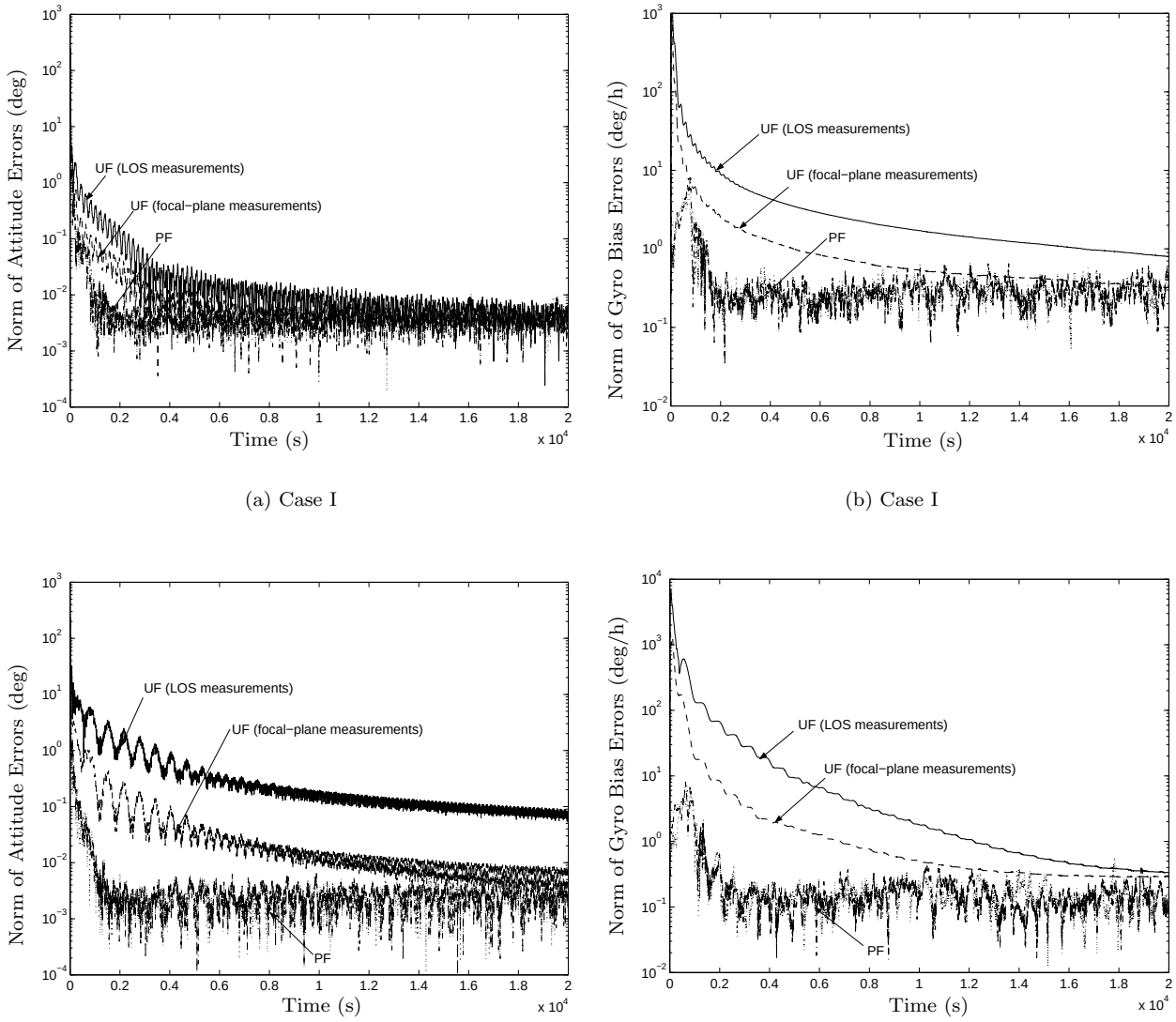
Figure 2. Convergence Properties of the BF

measurement model. The same number of particles, $\sigma_{\mathrm{ST}}(t)$, G, and resampling/roughening schedule are used for the gyroless setting. For simplicity, in the simulation tests we have assumed that the inertia matrix in the BF is identical to the true value and that the disturbance torque is the only error source in the dynamical model. (In the real world, however, model uncertainty is a significant problem with the dynamical model.) In each gyroless run, the true values for the initial attitude and initial attitude rate are respectively drawn from the uniform attitude distribution and a zero mean Gaussian distribution with the standard deviation of 1 deg/s per axis. The initial distributions of the BF are taken to be the same as the true distributions. In the next 200 runs, it is again observed that the BF always converges, whether at most 5 stars or a single star is used for update. The result of a typical run is given in Figure 2, in which the BF converges from the attitude error of 180° and the attitude rate error of about 0.04 rad/s.

Since the BF itself is based on Monte Carlo simulation, which is random in nature, and we cannot test all the initial conditions, we cannot assert from the previous 400 tests that the BF will never diverge. But the simulation results show that the BF does have a very high probability of convergence from very large attitude errors and very large gyro bias errors or attitude rate errors. Besides, it is worth pointing out that the same number of particles (i.e., 2000), the same G, and the same $\sigma_{\mathrm{ST}}(t)$ can be used for all these test runs.

The same simulation conditions are used for the tests of the UF. Unfortunately, the UF does not have as good convergence properties as the BF when the initial errors are extremely large. Many divergent runs are observed. When the UF does converge, it usually converges notably slower than the BF. This can be seen from the result of a typical test run for the star camera/gyros setting, given by Figure 3. The initial attitude error is about 150°; the initial gyro bias error is within 1 deg/hr per axis; and at most one star is used in the filters. The only difference between Case I and Case II of the typical run is in the attitude rate. In Case I, the initial attitude rate is $[0.03, 0.03, 0.1]^T$ rad/s; in Case II, the initial attitude rate is $[0.1, 0.03, 0.03]^T$ rad/s. The BF (with R-factor = 5 and $G = 0.12$) converges in about 2000 seconds in both cases, and both the attitude and gyro bias estimates have almost the same convergence rate. The UF does not fully converge at the end of the simulation, or 20000 seconds. The gyro bias estimates converges slower than the attitude estimates. In both cases the UF yields worse gyro bias estimates than the BF. The UF parameters α, β, and κ, which are used to control the relative positions and weights of the sigma points, now have a significant influence on the convergence properties of the UF. Inappropriate choice of α, β, and κ can result in filter divergence. Also, it is difficult to obtain overall good parameters by trial and error, if not impossible. In this typical run the UF parameters are tuned so as to fit the first case, but the same UF parameters are used in both cases. Clearly, when the initial attitude rate changes from $[0.03, 0.03, 0.1]^T$ rad/s (Case I) to $[0.1, 0.03, 0.03]^T$ rad/s (Case II), the performance of the UF degrades. If the initial attitude rate changes to $[0.03, 0.03, 0.03]^T$ rad/s, we observe the divergence of the UF with the LOS measurement model.

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337



(a) Case I

(b) Case I

(c) Case II

(d) Case II

Figure 3. Comparisons of Convergence Properties between the BF and the UF

If the initial attitude rate changes to $[0.02, 0.02, 0.02]^T$ rad/s, we observe the divergence of the UF with the focal-plane measurement model. Unlike the BF, which yields very similar results when used with different star camera measurement models (in Figure 3, the results of the BF with different measurement models cannot be discriminated), the UF behaves differently when different measurement models are used. Figure 3 shows that the UF with the focal-plane measurement model converges faster than the UF with the LOS measurement model in this specific run. The difference between the two measurement models becomes more significant when the initial attitude error is very close to 180°. It is observed that the UF with the LOS measurement model has more tendency to converge than the UF with the focal-plane model in the case of close to 180° attitude errors. This might be due to lower nonlinearity of the LOS measurement model and boundedness of the LOS measurements. To sum up, the UF does not always converge when the initial errors are extremely large and it usually has a lower rate of convergence than the BF. Although the UF parameters do not need to be tuned when the filter works in the linear regime, to tune these parameters in order to achieve overall good convergence properties is not easy. On the contrary, the BF requires much less parameter tuning and is insensitive to the variations of the attitude trajectories.

Downloaded by "Embry-Riddle Aeronautical University, Daytona Beach" on February 4, 2025 | http://arc.aiaa.org | DOI: 10.2514/6.2004-5337

VI. Conclusions

In this paper a new approach to attitude estimation using the Bootstrap Filter formulation was derived. Simulation results indicate that when the initial errors are small, the steady-state accuracy of the Bootstrap Filter may be worse than that of the Unscented Filter, because the Bootstrap Filter introduces extra Monte Carlo variations. However, for large initial errors the Bootstrap Filter has much better convergence properties than the Unscented Filter, because the Bootstrap Filter implicitly captures higher order characteristics of the posterior distribution. Therefore, it may be desirable to use the Bootstrap Filter in the initialization phase in order to better handle the problem of converging from extremely large initial errors. After convergence of the Bootstrap Filter, conventional filters, such the Unscented Filter or the Extended Kalman Filter, may be employed for the best possible steady-state accuracy. Though the Bootstrap Filter in this paper is proposed for star camera based attitude estimation, it can be easily extended to applications such as three-axis magnetometer based attitude estimation or GPS attitude estimation.

References

[1]Lefferts, E. J., Markley, F. L., and Shuster, M. D., "Kalman Filtering for Spacecraft Attitude Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 5, No. 5, Sept.-Oct. 1982, pp. 417–429.

[2]Crassidis, J. L. and Markley, F. L., "Unscented Filtering for Spacecraft Attitude Estimation," *Journal of Guidance, Control and Dynamics*, Vol. 26, No. 4, July-Aug. 2003, pp. 536–542.

[3]Psiaki, M. L., "Attitude-Determination Filtering via Extended Quaternion Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 2, March-April 2000, pp. 206–214.

[4]Kasdin, N. J., "Satellite Quaternion Estimation from Vector Measurements with the Two-Step Optimal Estimator," *AAS Guidance and Control Conference*, Breckenridge, CO, Feb. 2002, AAS 02-002.

[5]Shuster, M. D., "A Survey of Attitude Representations," *Journal of the Astronautical Sciences*, Vol. 41, No. 4, Oct.-Dec. 1993, pp. 439–517.

[6]Markley, F. L., "Attitude Representations for Kalman Filtering," *AAS/AIAA Astrodynamics Specialist Conference*, Quebec City, Quebec, Aug. 2001, AAS 01-309.

[7]Schaub, H. and Junkins, J. L., "Stereographic Orientation Parameters for Attitude Dynamics: A Generalization of the Rodrigues Parameters," *Journal of the Astronautical Sciences*, Vol. 44, No. 1, Jan.-March 1996, pp. 1–20.

[8]Gordon, N. J., Salmond, D. J., and Smith, A. F. M., "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F Vol. 140 No. 2*, Seattle, WA, April 1993, pp. 107–113.

[9]Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T., "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, February 2002, pp. 174–185.

[10]Wan, E. and van der Merwe, R., "The Unscented Kalman Filter," *Kalman Filtering and Neural Networks*, edited by S. Haykin, chap. 7, John Wiley & Sons, New York, NY, 2001.

[11]Murphy, K. and Russell, S., "Rao-Blackwellized Particle Filtering for Dynamic Bayesian Networks," *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon, chap. 24, Springer-Verlag, New York, NY, 2001.

[12]Deutscher, J., Davison, A., and Reid, I., "Articulated Body Motion Capture by Annealed Particle Filtering," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition Vol. 2*, Hilton Head Island, South Carolina, June 2000, pp. 126–133.

[13]Musso, C., Oudjane, N., and Gland, F. L., "Improving Regularized Particle Filters," *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon, chap. 12, Springer-Verlag, New York, NY, 2001.

[14]Liu, J. S. and Chen, R., "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, Vol. 93, No. 443, 1998, pp. 1032–1044.

[15]Markley, F. L., "Attitude Dynamics," *Spacecraft Attitude Determination and Control*, edited by J. R. Wertz, Kluwer Academic Publishers, The Netherlands, 1978.

[16]Farrenkopf, R. L., "Analytic Steady-State Accuracy Solutions for Two Common Spacecraft Attitude Estimators," *Journal of Guidance and Control*, Vol. 1, No. 4, July-Aug. 1978, pp. 282–284.

[17]Shuster, M. D., "Maximum Likelihood Estimation of Spacecraft Attitude," *The Journal of the Astronautical Sciences*, Vol. 37, No. 1, Jan.-March 1989, pp. 79–88.

[18]Shuster, M. D., "Uniform Attitude Probability Distributions," *to appear in Journal of the Astronautical Sciences*, Vol. 51, No. 4, 2003.

[19]Crassidis, J. L. and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 310–316.