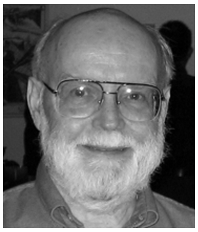
JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 30, No. 1, January–February 2007

Survey of Nonlinear Attitude Estimation Methods

John L. Crassidis
University at Buffalo, State University of New York, Amherst, New York 14260-4400
F. Landis Markley
NASA Goddard Space Flight Center, Greenbelt, Maryland 20771
and
Yang Cheng
University at Buffalo, State University of New York, Amherst, New York 14260-4400

DOI: 10.2514/1.22452

John L. Crassidis is an Associate Professor of Mechanical and Aerospace Engineering at the University at Buffalo (UB), State University of New York. He received his B.S., M.S., and Ph.D. in Mechanical Engineering from the State University of New York at Buffalo. Prior to joining UB in 2001, he held previous academic appointments at Catholic University of America from 1996 to 1998 and Texas A&M University from 1998 to 2001. From 1996 to 1998, he was a NASA Postdoctoral Research Fellow at Goddard Space Flight Center, where he worked on a number of spacecraft projects and research ventures involving attitude determination and control systems. He is the principal author of the textbook *Optimal Estimation of Dynamic Systems* and has authored or coauthored more than 100 journal and refereed conference papers. He has received many awards for his achievements including the best paper award for both the 2001 and 2003 AIAA Guidance, Navigation, and Control conferences, the 2006 AIAA Sustained Service Award, and the Society of Automotive Engineers 2006 Ralph R. Teetor Educational Award. His current research interests include nonlinear estimation and control theory, spacecraft attitude determination and control, attitude dynamics and kinematics, and robust vibration suppression. Since 1997 he has been a member of the AIAA Technical Committee on Guidance, Navigation, and Control where he currently serves as Chair. He is an Associate Fellow of AIAA.



F. Landis Markley received a Bachelor of Engineering Physics degree from Cornell University in 1962 and a Ph.D. in Physics from the University of California at Berkeley in 1967. While a member of the technical staff at the Computer Sciences Corporation from 1974 to 1978 he made major contributions to the essential book *Spacecraft Attitude Determination and Control*, edited by James R. Wertz, and directed a team that developed a FORTRAN prototype of the onboard attitude determination and control software for the Solar Maximum Mission spacecraft. He was a research physicist at the United States Naval Research Laboratory from 1978 to 1985, during which time he performed research in autonomous navigation and coauthored a widely cited Journal of Guidance, Control, and Dynamics survey article on Kalman filtering for spacecraft attitude determination, which was presented as an invited paper at the 20th Aerospace Sciences Meeting of the AIAA and was translated into Russian for publication in the Soviet journal *Aeronautics/Space Technology*. He has been employed at Goddard Space Flight Center since 1985. He headed a team that developed the algorithms and the software implementation of the general-purpose Attitude Determination Error Analysis System, which is used to develop specifications for attitude determination systems, and presented a paper on these error analysis methods at the international symposium on Space Dynamics at CNES, Toulouse, France, in October 1989. He has played a key role in developing attitude determination and control systems for several Goddard missions, including SAMPEX, GOES, TRMM, HST, and WMAP. He received the NASA Exceptional Service Medal in 1994 and again in 2005 and was appointed a Goddard Senior Fellow in 2000. He is a Fellow of AIAA and received their Mechanics and Control of Flight Award in 1998. He is a member of the American Astronautical Society and received their Dirk Brouwer Award in 2005.

Yang Cheng is a Research Assistant Professor of Mechanical and Aerospace Engineering at the University at Buffalo (UB), State University of New York. He received his B.Eng., M.Eng., and Ph.D. in aerospace engineering from the Harbin Institute of Technology, Harbin, China, in 1997, 1999, and 2003, respectively. From October 2003 to September 2006, he was a Postdoctoral Research Fellow at UB where he performed research on spacecraft attitude estimation, ground target tracking, data assimilation, and signal sampling and reconstruction. His research interests are in spacecraft attitude estimation, relative navigation, sensor calibration, nonlinear and robust estimation, and information fusion. He is a member of AIAA.

Received 13 January 2006; revision received 13 June 2006; accepted for publication 3 July 2006.

Nomenclature

A	=	attitude matrix
$\hat{A}$	=	estimated attitude matrix
$A(\boldsymbol{q})$	=	attitude matrix parameterized using the quaternion
$A_R(\boldsymbol{q})$	=	ray representation of the attitude matrix
$\boldsymbol{b}$	=	body-frame vector
$\tilde{\boldsymbol{b}}$	=	body-frame vector measurement
$\boldsymbol{d}$	=	model-error vector
$E\{\cdot\}$	=	expectation
$\boldsymbol{e}$	=	Euler axis
$\boldsymbol{F}$	=	external disturbance vector
$\boldsymbol{h}(\boldsymbol{x})$	=	nonlinear measurement model
$I_{3\times3}$	=	3×3 identity matrix
$\mathcal{J}$	=	inertial matrix
(N_v, N_u)	=	zero-mean Gaussian white-noise processes
P	=	covariance
$\mathcal{P}$	=	covariance for first step of two-step estimator
$(P^{\upsilon\upsilon}, P^{yy}, P^{xy})$	=	innovations, output, and cross-correlation covariances
p	=	probability density function
$\boldsymbol{p}$	=	vector of modified Rodrigues parameters
Q	=	process noise covariance
$\hat{\boldsymbol{q}}$	=	estimated quaternion
$\boldsymbol{q}$	=	quaternion
q_4	=	scalar part of the quaternion
R	=	measurement noise covariance
$(R_{ww}, R_{qq}, R_{xq}, R_{xx})$	=	weighting matrices in extended quaternion estimator
$\boldsymbol{r}$	=	reference-frame vector
$\boldsymbol{u}$	=	general control input vector
V	=	Lyapunov function
$\boldsymbol{v}$	=	measurement noise vector
$\boldsymbol{w}$	=	process noise vector
$\boldsymbol{w}_d$	=	deterministic dynamic disturbance vector
$\boldsymbol{w}_y$	=	deterministic measurement disturbance vector
$\boldsymbol{x}$	=	state vector
$\hat{\boldsymbol{x}}$	=	state estimate vector
$\mathcal{Y}$	=	first-step state in two-step estimator
$\hat{\mathcal{Y}}$	=	first-step estimate in two-step estimator
$\tilde{\boldsymbol{y}}$	=	measurement vector
$\boldsymbol{\beta}$	=	angular rate drift vector
Δt	=	sampling interval
$\Delta\boldsymbol{q}$	=	difference between true and estimated quaternion
$\Delta\boldsymbol{\beta}$	=	difference between true and estimated drift
$\boldsymbol{\delta}\hat{\boldsymbol{p}}$	=	estimated error generalized Rodrigues parameter vector
δq_4	=	scalar part of error quaternion
$\boldsymbol{\delta q}(\boldsymbol{\phi})$	=	multiplicative error quaternion
$\delta(t-\tau)$	=	Dirac-delta function
$\boldsymbol{\delta\varrho}$	=	vector part of error quaternion
$(\boldsymbol{\eta}_v, \boldsymbol{\eta}_u)$	=	gyro noise vectors
ϑ	=	rotation angle
$\boldsymbol{\varrho}$	=	vector part of the quaternion
σ	=	standard deviation
$\boldsymbol{\sigma}$	=	sigma points
(σ_v^2, σ_u^2)	=	gyro noise spectral densities
$\boldsymbol{\upsilon}$	=	noise for unit-vector measurement
$\Phi_{3\times3}$	=	attitude state transition matrix
$\Phi_{4\times4}$	=	quaternion state transition matrix
$\boldsymbol{\phi}$	=	vector for error quaternion
$\tilde{\boldsymbol{\omega}}$	=	measured angular rate vector
$\boldsymbol{\omega}$	=	angular rate vector
$0_{3\times3}$	=	3×3 matrix of zeros

I. Introduction

ATTITUDE estimation involves a two-part process: 1) estimation of a vehicle's orientation from body measurements and known reference observations, such as line-of-sight measurements to known observed stars, and 2) filtering of noisy measurements. The second part is achieved by combining the measurements with models, which in itself can be done a number of different ways. One way is to use a kinematics model propagated with three-axis rate-integrating gyros. However, the rates measured by gyros drift over time. Therefore, the attitude state vector is usually appended by three more states to determine this drift. This leads to a complementary approach, in which the gyros are used to filter the noisy body measurements and the measurements are used to determine the drift inherent in the gyros. Another way involves combining the kinematics model with a dynamics model for the angular rate. However, even a detailed dynamics model, such as Euler's rotational equations, will have inherent errors. For example, the inertia matrix may not be well known. This is compensated in filter designs by using process noise, which leads to the classic "tuning" problem in the filter. Throughout this paper the terms "filter" and "estimator" are used synonymously, because noisy measurements are involved. When perfect observations are given, then the term "observer" is used. Another term is "smoother," which refers to a batch algorithm, that is, not executed in real time, to provide better estimates than a real-time filtering algorithm. Various attitude filters, observers and smoothers are shown here.

The extended Kalman filter [1–7] (EKF) is the workhorse of real-time spacecraft attitude estimation. Because SO(3) (the special orthogonal group for three-dimensional space) has dimension three, most attitude determination EKFs use lower-dimensional attitude parameterizations than the nine-parameter attitude matrix itself. The fact that all three-parameter representations of SO(3) are singular or discontinuous for certain attitudes [8] has led to extended discussions of constraints and attitude representations in EKFs [6,7,9–11]. These issues are now well understood, however, and the EKF, especially in the form known as the multiplicative extended Kalman filter [5–7] (MEKF), has performed admirably in the vast majority of attitude determination applications. Nevertheless, poor performance or even divergence arising from the linearization implicit in the EKF has led to the development of other filters. Several of these approaches retain the basic structure of the EKF, such as additive EKF approaches [12–17], a backwards-smoothing EKF [18], and deterministic EKF-like estimators [19–21], which are closely related to H_∞ control design [22]. Another approach applies a two-step optimal estimator, in which the first is a transformation of the measurement equation to a linear one, thus providing a linear filter design, and the second step is a least-squares solution for the estimate [23].

Other designs use various assumptions to derive simplified filters. These generally provide suboptimal performance characteristics in relation to the EKF, but involve linear or pseudolinear equations that are used to estimate the states of a nonlinear dynamical system. Therefore, linear design and analysis tools can be used to construct the filter and assess its overall performance. Some of these use a point-by-point solution of the attitude, for example, methods that are based on the quaternion estimator (QUEST) attitude determination solution [24]. Simple filter designs based on QUEST include filter QUEST [25] and recursive QUEST [26]. A more complicated but far more robust approach, called extended QUEST [27], uses a full nonlinear propagation along with a novel measurement update. This approach can be used to estimate attitude and additional parameters as well.

Several new alternatives to the standard EKF have been recently introduced, such as unscented filters (UFs) [28–31], also known as sigma-point filters, and particle filters [32–35]. Unscented filters are essentially based on second or higher-order approximations of nonlinear functions, which are used to estimate the mean and covariance of the state vector. Though the mean and covariance are sufficient to represent a Gaussian distribution, they are not sufficient to represent a general probability distribution. This may be overcome by using particle filters [32–35] (PFs). A PF for attitude estimation based on the bootstrap filter [33] is shown in [34]. Oshman and Carmi [35] use a bootstrap filter to estimate the quaternion and employ a genetic algorithm to estimate the gyro bias parameters, avoiding the need to augment the particle filter's state and thus reducing the number of particles required. Another approach uses a higher-order Taylor series expansion than the standard EKF to predict the estimate at the current time, based on current-time measurements and previous-time estimates [36].

The optimal solution of the nonlinear estimation problem requires the propagation of the conditional probability density function (PDF) of the state given the observation history [35]. All practical nonlinear filters are approximations to this ideal. Exact finite dimensional filters [37] can be found that solve some nonlinear problems by using the Fokker–Planck equation [2,38] to propagate a non-Gaussian PDF between measurements and Bayes' formula [1,2] to incorporate measurement information. A recently proposed filter [39] follows this pattern, but does not solve the nonlinear attitude filtering problem exactly. This is referred to as an orthogonal filter, because it represents the attitude by an orthogonal rotation matrix, rather than by some parameterization of the rotation matrix. Nonlinear observers often exhibit global convergence, which is to say that they can converge from any initial guess, as discussed in the survey paper of [40]. More recent observers are high gain observers [41], normal form observers [42], backstepping observers [43], observers based on a moving horizon [44,45], and pseudospectral observers [46]. Several applications of observers for attitude control have been proposed [47–52].

Adaptive approaches generally fall into two categories. One category encompasses approaches that adaptively tune the Kalman filter through the identification of either the process noise covariance or measurement noise covariance, or both simultaneously [53]. An example is the approach demonstrated in [54] that uses linearized equations. Another approach develops adaptive filters that address both colored and white-noise statistics [55]. An adaptive filter is also proposed in [17] to account for inaccuracy in the knowledge of the process noise statistical model, which uses a linear pseudomeasurement model. Other adaptive approaches use adaptive methods for fault tolerant estimation purposes [56,57]. The other category includes approaches that adaptively estimate unknown system parameters, such as the inertia matrix. These generally fall into two basic categories: 1) parameter estimation or filter-based methods, and 2) nonlinear adaptive techniques. Least-squares methods to determine the inertia matrix and other constant parameters, such as disturbance model parameters and biases, are shown in [58–60]. A disturbance accommodation technique that models the unknown disturbance angular rate using polynomials in time as basis functions is shown in [61]. Nonlinear adaptive techniques are similar to nonlinear observers in that they usually provide global stability proofs that guarantee convergence of the estimated parameters [62–64].

This paper will review the basic assumptions of these filters, observers, and smoothers, presenting enough mathematical detail to give a general orientation. First, reviews of the quaternion parameterization and gyro model equations are given. Then, attitude estimation methods based on the EKF are shown, followed by QUEST-based approaches. Next, the two-step estimator is shown. The UF and PF approaches are then shown, followed by the orthogonal filter. Then, the predictive filter, as well as nonlinear observers and adaptive approaches are reviewed. The paper concludes with a discussion of the strengths and weaknesses of the various filters.

II. Quaternion Parameterization and Gyro Model

The attitude of a vehicle is defined as its orientation with respect to some reference frame. If the reference frame is nonmoving, then it is commonly referred to as an *inertial* frame. To describe the attitude two coordinate systems are usually defined: one on the vehicle body and one on the reference frame. For most dynamical applications these coordinate systems have orthogonal unit vectors that follow the right-hand rule. The attitude matrix A, often referred to as the direction cosine matrix or rotation matrix, maps one frame to another. The attitude matrix is an orthogonal matrix, that is, its inverse is given by its transpose, and proper, that is, its determinant is $+1$ [65]. For spacecraft applications the attitude mapping is usually applied from the reference frame to the vehicle body frame. Mathematically, the mapping from the reference frame to the body frame is given by

$$\boldsymbol{b} = A\boldsymbol{r} \tag{1}$$

where $\boldsymbol{b}$ is the body-frame vector and $\boldsymbol{r}$ is the reference-frame vector.

Several parameterizations of the attitude are possible [66]. Minimal parameterizations, such as the Euler angles, the Rodrigues parameters (Gibb's vector) and the modified Rodrigues parameters (MRPs), are often avoided in filter designs for the "global attitude" due to their associated singularities [8]. They are often used to define the "local error attitude" though, which is discussed later. For modern-day applications, that is, since the early 1980s, the quaternion [67] has been the most widely used attitude parameterization. The quaternion is a four-dimensional vector, defined as

$$\boldsymbol{q} \equiv \begin{bmatrix} \boldsymbol{\varrho} \\ q_4 \end{bmatrix} \tag{2}$$

with

$$\boldsymbol{\varrho} \equiv \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}^T = \boldsymbol{e}\sin(\vartheta/2) \tag{3a}$$

$$q_4 = \cos(\vartheta/2) \tag{3b}$$

where $\boldsymbol{e}$ is the unit Euler axis and ϑ is the rotation angle. Because a four-dimensional vector is used to describe three dimensions, the quaternion components cannot be independent of each other. The quaternion satisfies a single constraint given by $\boldsymbol{q}^T\boldsymbol{q} = 1$. The attitude matrix is related to the quaternion by

$$A(\boldsymbol{q}) = (q_4^2 - \|\boldsymbol{\varrho}\|^2)I_{3\times 3} + 2\boldsymbol{\varrho}\boldsymbol{\varrho}^T - 2q_4[\boldsymbol{\varrho}\times] = \Xi^T(\boldsymbol{q})\Psi(\boldsymbol{q}) \tag{4}$$

where $I_{3\times 3}$ is a 3×3 identity matrix and

$$\Xi(\boldsymbol{q}) \equiv \begin{bmatrix} q_4 I_{3\times 3} + [\boldsymbol{\varrho}\times] \\ -\boldsymbol{\varrho}^T \end{bmatrix} \tag{5a}$$

$$\Psi(\boldsymbol{q}) \equiv \begin{bmatrix} q_4 I_{3\times 3} - [\boldsymbol{\varrho}\times] \\ -\boldsymbol{\varrho}^T \end{bmatrix} \tag{5b}$$

Also, $[\boldsymbol{\varrho}\times]$ is the cross-product matrix defined by

$$[\boldsymbol{\varrho}\times] \equiv \begin{bmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{bmatrix} \tag{6}$$

For small angles the vector part of the quaternion is approximately equal to half angles [66], which will be used later.

The quaternion kinematics equation is given by

$$\dot{\boldsymbol{q}} = \tfrac{1}{2}\Xi(\boldsymbol{q})\boldsymbol{\omega} = \tfrac{1}{2}\Omega(\boldsymbol{\omega})\boldsymbol{q} \tag{7}$$

where $\boldsymbol{\omega}$ is the three-component angular rate vector and

$$\Omega(\boldsymbol{\omega}) \equiv \begin{bmatrix} -[\boldsymbol{\omega}\times] & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix} \tag{8}$$

A useful identity is given by

$$\Psi(\boldsymbol{q})\boldsymbol{\omega} = \Gamma(\boldsymbol{\omega})\boldsymbol{q} \tag{9}$$

where

$$\Gamma(\boldsymbol{\omega}) \equiv \begin{bmatrix} [\boldsymbol{\omega}\times] & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix} \tag{10}$$

A major advantage of using the quaternion is that the kinematics equation is linear in the quaternion and is also free of singularities. Another advantage of the quaternion is that successive rotations can be accomplished using quaternion multiplication. Here the convention of [6] is adopted, where the quaternions are multiplied in the same order as the attitude matrix multiplication, in contrast to the usual convention established by Hamilton [67]. A successive rotation is written using

$$A(\boldsymbol{q}')A(\boldsymbol{q}) = A(\boldsymbol{q}' \otimes \boldsymbol{q}) \tag{11}$$

The composition of the quaternions is bilinear, with

$$\boldsymbol{q}' \otimes \boldsymbol{q} = \begin{bmatrix} \Psi(\boldsymbol{q}') & \boldsymbol{q}' \end{bmatrix}\boldsymbol{q} = \begin{bmatrix} \Xi(\boldsymbol{q}) & \boldsymbol{q} \end{bmatrix}\boldsymbol{q}' \tag{12}$$

Also, the inverse quaternion is defined by

$$\boldsymbol{q}^{-1} \equiv \begin{bmatrix} -\boldsymbol{\varrho} \\ q_4 \end{bmatrix} \tag{13}$$

Note that $\boldsymbol{q} \otimes \boldsymbol{q}^{-1} = [0\ 0\ 0\ 1]^T$, which is the identity quaternion. A computationally efficient algorithm to extract the quaternion from the attitude matrix is given in [68]. A more thorough review of the quaternion parameterization, as well as other parameterizations, can be found in the survey paper by Shuster [66] and in the book by Kuipers [69].

A common sensor that measures the angular rate is a rate-integrating gyro. For this sensor, a widely used three-axis continuous-time model is given by [70]

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \boldsymbol{\beta} + \boldsymbol{\eta}_v \tag{14a}$$

$$\dot{\boldsymbol{\beta}} = \boldsymbol{\eta}_u \tag{14b}$$

where $\tilde{\boldsymbol{\omega}}$ is the measured rate, $\boldsymbol{\beta}$ is the drift, and $\boldsymbol{\eta}_v$ and $\boldsymbol{\eta}_u$ are independent zero-mean Gaussian white-noise processes with

$$E\{\boldsymbol{\eta}_v(t)\boldsymbol{\eta}_v^T(\tau)\} = \sigma_v^2\delta(t-\tau)I_{3\times 3} \tag{15a}$$

$$E\{\boldsymbol{\eta}_u(t)\boldsymbol{\eta}_u^T(\tau)\} = \sigma_u^2\delta(t-\tau)I_{3\times 3} \tag{15b}$$

where $E\{\cdot\}$ denotes expectation and $\delta(t-\tau)$ is the Dirac-delta function. A more general gyro model includes scale factors and misalignments, which can also be estimated in real time [71,72]. For simulation purposes, discrete-time gyro measurements can be generated using the following equations [73]:

$$\tilde{\boldsymbol{\omega}}_{k+1} = \boldsymbol{\omega}_{k+1} + \frac{1}{2}[\boldsymbol{\beta}_{k+1} + \boldsymbol{\beta}_k] + \left[\frac{\sigma_v^2}{\Delta t} + \frac{1}{12}\sigma_u^2\Delta t\right]^{1/2}\boldsymbol{N}_v \tag{16a}$$

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k + \sigma_u \Delta t^{1/2}\boldsymbol{N}_u \tag{16b}$$

where the subscript k denotes the kth time step, Δt is the gyro sampling interval, and $\boldsymbol{N}_v$ and $\boldsymbol{N}_u$ are zero-mean Gaussian white-noise processes with covariance each given by the identity matrix.

III. Extended Kalman Filter

The most straightforward way to attack a nonlinear estimation problem is to linearize about the current best estimate. This leads, of course, to the EKF [3], which is the workhorse of satellite attitude determination. There are several different implementations of the attitude EKF, depending on both the attitude representation [66] used in the state vector and the form in which observations are input. It is a well-known fact that all globally continuous and nonsingular representations of the rotations have at least one redundant component [8], leading to alternatives using an attitude representation that is either singular or redundant. These alternatives can be divided into three general classes, which are referred to as the minimal representation extended Kalman filter, the multiplicative extended Kalman filter (MEKF), and the additive extended Kalman filter (AEKF). An overview of Kalman filtering for spacecraft attitude estimation emphasizing the quaternion representation, with a complete list of references through 1981 is shown in [6]. This section will provide a brief overview, emphasizing developments since 1981.

A. Observation Preprocessing in an EKF

Almost all attitude measurements can be converted to unit vectors: centroids in a star tracker's focal plane to star unit vectors, horizon sensor measurements to a nadir-pointing unit vector, or triaxial magnetometer measurements to a unit vector along the magnetic field, for example. It has proved convenient, therefore, to develop a standard unit-vector interface for the EKF, resulting in the unit-vector filter (UVF) [74,75]. The UVF also employs a very useful approximation to the errors in the unit-vector measurements. In particular, Shuster [24] has shown that nearly all the probability of the errors is concentrated on a very small area about the direction of $A\boldsymbol{r}$, and so the sphere containing that point can be approximated by a tangent plane, characterized by

$$\tilde{\boldsymbol{b}} = A\boldsymbol{r} + \boldsymbol{\upsilon}, \qquad \boldsymbol{\upsilon}^T A\boldsymbol{r} = 0 \tag{17}$$

where $\tilde{\boldsymbol{b}}$ denotes the measurement and the sensor error $\boldsymbol{\upsilon}$ is approximately Gaussian, which satisfies

$$E\{\boldsymbol{\upsilon}\} = \boldsymbol{0} \tag{18a}$$

$$R \equiv E\{\boldsymbol{\upsilon}\boldsymbol{\upsilon}^T\} = \sigma^2[I_{3\times 3} - (A\boldsymbol{r})(A\boldsymbol{r})^T] \tag{18b}$$

Equation (18b) is known as the QUEST measurement model [24,74,75]. This model is quite accurate for small field-of-view sensors. The approximations in this error model are discussed in [11,76]. Equation (18b) gives a rank-deficient R matrix, which would appear to give rise to problems for the EKF, but Shuster has shown that the simpler, full-rank form

$$R = \sigma^2 I_{3\times 3} \tag{19}$$

gives equivalent results in this context [74,77]. The QUEST measurement model has been expanded for large field-of-views in [76].

Using QUEST [24] or an equivalent quaternion estimator as a data compressor simplifies the interface even further. This is especially useful because many modern star trackers compute a quaternion from multiple star vectors, and the quaternion output from the star tracker provides a convenient "measurement" for input to an EKF [7,74,78].

B. Minimal Representation Extended Kalman Filter

The rotation group has three dimensions, and so the most straightforward implementation of an EKF employs a three-dimensional parameterization of the attitude. The earliest known published attitude EKF used the 1-2-3 sequence of Euler angles [4]. It is well known that these angles have a "gimbal lock" singularity when the magnitude of the middle angle is 90 deg, and so this form of the EKF is most appropriate when the spacecraft does not stray too far from a reference attitude. A good example is an Earth-pointing spacecraft, with the attitude being defined with respect to a local-vertical/local-horizontal coordinate frame. If the gimbal lock condition arises, the coordinate axes to which the vehicle attitude is referenced must be repeatedly shifted to avoid singularity [4].

Euler angles are inappropriate for agile spacecraft, such as astronomical observatories. Minimal representation EKFs employing the Rodrigues parameters [79] and the MRPs have been developed for this application [80]. The MRPs are nonsingular for rotations less than 360 deg, and the singularity can be avoided by changing to a "shadow set" of parameters [81]. This form of the EKF has not found wide application, however.

C. Multiplicative EKF

The MEKF represents the attitude as the product of an estimated attitude and a deviation from that estimate. A nonsingular representation of the estimated attitude and a three-parameter representation of the deviation are employed. The most usual implementation uses the quaternion representation for the attitude [5–7,82–84]. In this case the product is

$$\boldsymbol{q} = \boldsymbol{\delta q}(\boldsymbol{\phi}) \otimes \hat{\boldsymbol{q}} \tag{20}$$

where $\hat{\boldsymbol{q}}$ is the unit estimated quaternion and $\boldsymbol{\delta q}(\boldsymbol{\phi})$ is a unit quaternion representing the rotation from $\hat{\boldsymbol{q}}$ to the true attitude $\boldsymbol{q}$, parameterized by a three-component vector $\boldsymbol{\phi}$.

An alternative formulation, which has some advantages, reverses the order of multiplication in Eq. (20) so that $\boldsymbol{\phi}$ represents the attitude errors in the inertial reference frame rather than in the body frame [85–87]. It is also possible to represent the reference attitude by an estimated attitude matrix $\hat{A}$ rather than by an estimated quaternion [7,85,88,89]. This requires more parameters, but may save computations if the attitude matrix is explicitly required. An argument in favor of the quaternion is that it is easy to restore normalization that may be lost due to numerical errors, whereas restoring the orthogonality of $\hat{A}$ is nontrivial. Gray has argued that this argument is not compelling if reasonable computational care is taken [85].

The representation of Eq. (20) is clearly redundant. The basic idea of the MEKF is that the EKF estimates the three-vector $\boldsymbol{\phi}$, whereas the correctly normalized four-component $\hat{\boldsymbol{q}}$ provides a globally nonsingular attitude representation. If $\hat{\boldsymbol{\phi}} \equiv E\{\boldsymbol{\phi}\}$ is the estimate of $\boldsymbol{\phi}$, then Eq. (20) says that $\boldsymbol{\delta q}(\hat{\boldsymbol{\phi}}) \otimes \hat{\boldsymbol{q}}$ is the estimate of the true attitude quaternion $\boldsymbol{q}$. This is equal to $\hat{\boldsymbol{q}}$ if the redundancy in the attitude representation is removed by ensuring that $\boldsymbol{\phi}$ has zero mean so that $\boldsymbol{\delta q}(\hat{\boldsymbol{\phi}}) = \boldsymbol{\delta q}(\boldsymbol{0})$ is the identity quaternion. This choice means that $\boldsymbol{\phi}$ is a three-component representation of the attitude error and its covariance is the attitude error covariance in the body frame. The fundamental advantages of the MEKF are that $\hat{\boldsymbol{q}}$ is a unit quaternion by definition, the covariance matrix has the minimum dimensionality, and the three-vector $\boldsymbol{\phi}$ never approaches a singularity, because it represents only small attitude errors.

Several choices for $\boldsymbol{\phi}$ have been used [7], including the vector of infinitesimal rotation angles [84], 2 times the vector part of the quaternion [6], 2 times the vector of Rodrigues parameters [7], 4 times the vector of MRPs, or the integrated rate parameters [89]. All these choices have the small-angle approximation

$$\boldsymbol{\delta q}\,(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{\phi}/2 \\ 1 \end{bmatrix} + \mathcal{O}(\|\boldsymbol{\phi}\|^2) \tag{21}$$

and MEKFs employing them differ only in third order in the measurement updates to the error angle [7].

The MEKF was first used in the space precision attitude reference system (SPARS) in 1969 [82,88], was later developed for NASA's multimission modular spacecraft [5], and has been used for attitude estimation on board several NASA spacecraft. It has been discussed in detail in [6,7]. The latter reference discusses the extension of the MEKF to a second-order filter, following earlier work by Vathsal [90].

D. Additive Extended Kalman Filter

An AEKF uses a nonsingular parameterization of the attitude in the filter's state vector. Almost all AEKFs have employed the quaternion [12–17], but the attitude matrix itself has also been employed [91]. The AEKF can be either unconstrained or constrained so that the attitude matrix is orthogonal or, equivalently, that the quaternion has unit norm. A properly constrained AEKF is mathematically equivalent to the MEKF [6,9,10].

Only the unconstrained quaternion AEKF will be discussed here. This filter relaxes the quaternion normalization condition and treats the four components of the quaternion as independent parameters. It defines the estimate $\hat{\boldsymbol{q}}$ and error $\Delta\boldsymbol{q}$ by

$$\hat{\boldsymbol{q}} \equiv E\{\boldsymbol{q}|\tilde{\boldsymbol{y}}\} \quad \text{and} \quad \Delta\boldsymbol{q} \equiv \boldsymbol{q} - \hat{\boldsymbol{q}} \tag{22}$$

This means that

$$E\{\|\boldsymbol{q}\|^2|\tilde{\boldsymbol{y}}\} = E\{\|\hat{\boldsymbol{q}} + \Delta\boldsymbol{q}\|^2|\tilde{\boldsymbol{y}}\} = \|\hat{\boldsymbol{q}}\|^2 + E\{\|\Delta\boldsymbol{q}\|^2|\tilde{\boldsymbol{y}}\} \geq \|\hat{\boldsymbol{q}}\|^2 \tag{23}$$

where $\tilde{\boldsymbol{y}}$ denotes the measurement vector. The equality in Eq. (23) is valid only if $\Delta\boldsymbol{q}$ is identically zero. Equation (23) shows that if the random variable $\boldsymbol{q}$ has unit norm and is not error free, the norm of its expectation must be less than unity. The usual expression for the attitude matrix as a homogenous quadratic function of the quaternion gives an orthogonal matrix only if the quaternion has unit norm. Many unconstrained AEKFs have used the homogenous quadratic form, which means that the attitude matrix is only approximately orthogonal. It can be shown to approach orthogonality as the filter converges, however [12–14].

The attitude matrix is guaranteed to be orthogonal if it is computed using the normalized quaternion $\boldsymbol{q}/\|\boldsymbol{q}\|$ in the homogenous quadratic form, giving

$$A_R(\boldsymbol{q}) = \|\boldsymbol{q}\|^{-2}\Big\{\Big(q_4^2 - \|\boldsymbol{\varrho}\|^2\Big)I_{3\times3} + 2\boldsymbol{\varrho}\boldsymbol{\varrho}^T - 2q_4[\boldsymbol{\varrho}\times]\Big\} \tag{24}$$

The subscript R identifies this as the ray representation model, because any quaternion along a ray in Euclidean quaternion space (a straight line through the origin) represents the same attitude, with the exception of the zero quaternion at the origin. This is also known as the linearized orthogonalized matrix model [13,14]. The ray representation form of the unconstrained AEKF, which is effectively equivalent to the MEKF in the limit of continuous measurements [11], has been applied to attitude estimation of the array of low-energy x-ray imaging sensors (ALEXIS) and Cleft Accelerated Plasma Experimental Rocket (CAPER) spacecraft [15,16].

Choukroun et al. formulate a measurement model based on the matrix factorization of the attitude matrix [17] shown in Eq. (4). This gives a measurement model that is linear in the quaternion, but with state-dependent measurement noise

$$\begin{aligned} \tilde{\boldsymbol{y}} &= \Xi(\boldsymbol{q})(\boldsymbol{b} + \boldsymbol{v}) - \Psi(\boldsymbol{q})\boldsymbol{r} \\ &= \begin{bmatrix} -[(\boldsymbol{b}+\boldsymbol{r})\times] & (\boldsymbol{b}-\boldsymbol{r}) \\ -(\boldsymbol{b}-\boldsymbol{r})^T & 0 \end{bmatrix}\boldsymbol{q} + \Xi(\boldsymbol{q})\boldsymbol{v} \end{aligned} \tag{25}$$

where $\boldsymbol{v}$ is the measurement noise, which may or may not correspond to a unit-vector observation. This model gives the same covariance and state propagation as the ray representation AEKF, in the limit of continuous measurements [11], but has subtle differences for discrete measurements.

The relative merits of the AEKF and the MEKF have been discussed at length [9–11,92]. One disadvantage of the unconstrained AEKF is that its covariance matrix includes elements expressing the variance of the quaternion norm uncertainty and the correlation of the norm uncertainty with all other estimated parameters. These terms, which are not present in the MEKF, are neither conceptually nor computationally desirable.

E. Backwards-Smoothing Extended Kalman Filter

The kth step in a nonlinear filtering problem can be posed as a maximum a posteriori probability (MAP) estimation problem by writing the PDF as $p_k = \exp(-J_k)$ with the loss function

$$\begin{aligned} J_k = \frac{1}{2}\sum_{i=0}^{k-1}\Big\{&[\tilde{\boldsymbol{y}}_{i+1} - \boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})]^T R_{i+1}^{-1}[\tilde{\boldsymbol{y}}_{i+1} - \boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})] \\ &+ \boldsymbol{w}_i^T Q_i^{-1}\boldsymbol{w}_i\Big\} + \frac{1}{2}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^T P_0^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0) \end{aligned} \tag{26}$$

The MAP estimate $\hat{\boldsymbol{x}}_k$ is the vector $\boldsymbol{x}_k$ that, along with $\boldsymbol{x}_i$ and process noise $\boldsymbol{w}_i$ for $i = 0, 1, \ldots, k-1$, minimizes J_k subject to the dynamics equation

$$\boldsymbol{x}_{i+1} = \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{w}_i) \quad \text{for } i = 0, 1, \ldots, k-1 \tag{27}$$

The process noise covariance is Q_i, the measurement noise covariance is R_i, $\tilde{\boldsymbol{y}}_{i+1}$ is the measurement at time t_{i+1}, $\boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})$ is the nonlinear measurement model, and $\hat{\boldsymbol{x}}_0$ is the a priori estimate of the state with covariance P_0. It can be seen that the size of this problem grows with k. The usual EKF avoids this growth by not explicitly recomputing the values of $\hat{\boldsymbol{x}}_i$ for $i < k$ when $\boldsymbol{x}_k$ is optimized in the kth step. The iterated EKF improves upon the EKF by iterating the nonlinear measurement update equation for $\hat{\boldsymbol{x}}_k$, relinearizing about the updated state estimate at each iteration [1,93], but it does not explicitly recompute the values of $\hat{\boldsymbol{x}}_i$ for $i < k$. Any Kalman filter implicitly recomputes the past state estimates at a new measurement update; but this point is often overlooked because estimates in the past are generally of no interest. For linear dynamics and measurements, these past estimates are optimal, but they are not optimal with nonlinear dynamics or measurements. Thus the EKF linearizations of the past measurements and dynamics are not about the optimal estimates.

The backwards-smoothing extended Kalman filter (BSEKF), or superiterated EKF [18], improves on the iterated EKF by relinearizing a finite number of measurements in the past when a new measurement is processed. The BSEKF therefore combines some of the properties of an EKF, a smoother and a sliding-batch estimator. It finds $\boldsymbol{x}_k$ along with $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ for $i = k - m(k), \ldots, k - 1$ to minimize the loss function

$$\begin{aligned} J_k = \frac{1}{2} \sum_{i=k-m(k)}^{k-1} \Big\{ & [\tilde{\boldsymbol{y}}_{i+1} - \boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})]^T R_{i+1}^{-1} [\tilde{\boldsymbol{y}}_{i+1} - \boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})] \\ & + \boldsymbol{w}_i^T Q_i^{-1} \boldsymbol{w}_i \Big\} + \frac{1}{2} [\boldsymbol{x}_{k-m(k)} - \hat{\boldsymbol{x}}^*_{k-m(k)}]^T [P^*_{k-m(k)}]^{-1} [\boldsymbol{x}_{k-m(k)} \\ & - \hat{\boldsymbol{x}}^*_{k-m(k)}] \end{aligned} \tag{28}$$

where $\hat{\boldsymbol{x}}^*_{k-m(k)}$ is used to represent old data. Equation (28) is subject to the dynamics of Eq. (27) for $i \geq k - m(k)$. The loss function of Eq. (28) retains all of the nonlinearities of the most recent $m(k)$ stages, but the nonlinear effects of all the previous stages are represented by the quadratic second term, which is an approximation to the loss function $J_{k-m(k)}$ for fixed $\boldsymbol{x}_{k-m(k)}$ optimized over all the $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ for $i < k - m(k)$. A value m_{target} for the number of stages to be retained is chosen to balance accuracy and computational effort. When $k \leq m_{\text{target}}$, the BSEKF uses $m(k) = k$ stages, and when $k > m_{\text{target}}$, it uses $m(k) = m_{\text{target}}$ stages. The detailed steps of this procedure are presented in [18].

F. Deterministic EKF-Like Estimator

Nonlinear attitude estimators have been developed based on H_∞ control design techniques [19–21]. They have a structure similar to the EKF, but do not depend on the questionable assumption of white noise. One of these estimators has been applied to the attitude determination of the radar calibration (RADCAL) satellite using global positioning system (GPS) measurements [21].

The quaternion kinematic model is given by Eq. (7), but Eq. (14) is replaced by

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \boldsymbol{\beta} + B_1 \boldsymbol{w}_d \tag{29a}$$

$$\dot{\boldsymbol{\beta}} = B_2 \boldsymbol{w}_d \tag{29b}$$

where $\boldsymbol{w}_d$ is a square-integrable six-component deterministic dynamic disturbance vector, and B_1 and B_2 are known matrices. It is assumed that $B_1 B_2^T = 0_{3\times 3}$, although this assumption is not necessary. The kinematics can be written with $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{q}^T & \boldsymbol{\beta}^T \end{bmatrix}^T$ as

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + G(\boldsymbol{x}) \boldsymbol{w}_d \tag{30}$$

where

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \frac{1}{2} \Xi(\boldsymbol{q})(\tilde{\boldsymbol{\omega}} - \boldsymbol{\beta}) \\ 0_{3\times 1} \end{bmatrix} \quad \text{and} \quad G(\boldsymbol{x}) = \begin{bmatrix} -\frac{1}{2} \Xi(\boldsymbol{q}) B_1 \\ B_2 \end{bmatrix} \tag{31}$$

Continuous measurements are modeled as

$$\tilde{\boldsymbol{y}} = \boldsymbol{h}(\boldsymbol{x}) + D(\boldsymbol{x}) \boldsymbol{w}_y \tag{32}$$

where $\boldsymbol{h}(\boldsymbol{x})$ is the known m-component output vector, $\boldsymbol{w}_y$ is a square-integrable m-component deterministic measurement disturbance vector and D is a known matrix function of $\boldsymbol{x}$. The H_∞ estimation problem is defined as follows. Let

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{f}(\hat{\boldsymbol{x}}) + K(\hat{\boldsymbol{x}}, t)[\tilde{\boldsymbol{y}} - \boldsymbol{h}(\hat{\boldsymbol{x}})] \tag{33}$$

with initial value $\hat{\boldsymbol{x}}(0) = \hat{\boldsymbol{x}}_0$, and define an error function $\boldsymbol{z}$ by

$$\boldsymbol{z} = \boldsymbol{\zeta}(\boldsymbol{x}) - Z^T(\boldsymbol{x}) \hat{\boldsymbol{x}} \tag{34}$$

where all the functions and matrices are smooth functions with appropriate dimensions. Then the H_∞ estimation problem is to choose, for a given $\gamma > 0$ and a given metric N, the function $K(\hat{\boldsymbol{x}}, t)$ for which

$$\int_0^T \|\boldsymbol{z}(\tau)\|^2 \, \mathrm{d}\tau \leq \gamma^2 \left\{ N(\boldsymbol{x}_0, \hat{\boldsymbol{x}}_0) + \int_0^T [\|\boldsymbol{w}_d(\tau)\|^2 + \|\boldsymbol{w}_y(\tau)\|^2] \, \mathrm{d}\tau \right\} \tag{35}$$

Standard arguments show that a sufficient condition for the solution of this problem is the existence of a non-negative function $V(\boldsymbol{x}, \hat{\boldsymbol{x}}, t)$ with $V(\boldsymbol{x}_0, \hat{\boldsymbol{x}}_0, t) = \gamma^2 N(\boldsymbol{x}_0, \hat{\boldsymbol{x}}_0)$ that satisfies

$$\begin{aligned} J \equiv {} & \partial V/\partial t + (\partial V/\partial \boldsymbol{x}) \boldsymbol{f}(\boldsymbol{x}) + (\partial V/\partial \hat{\boldsymbol{x}}) \{ \boldsymbol{f}(\hat{\boldsymbol{x}}) + K(\hat{\boldsymbol{x}}, t)[\boldsymbol{h}(\boldsymbol{x}) \\ & - \boldsymbol{h}(\hat{\boldsymbol{x}})] \} + \boldsymbol{z}^T \boldsymbol{z} + \tfrac{1}{4} \gamma^{-2} [(\partial V/\partial \boldsymbol{x}) G(\boldsymbol{x}) G^T(\boldsymbol{x}) (\partial V/\partial \boldsymbol{x})^T \\ & + (\partial V/\partial \hat{\boldsymbol{x}}) K(\hat{\boldsymbol{x}}, t) R(\boldsymbol{x}) K^T(\hat{\boldsymbol{x}}, t) (\partial V/\partial \hat{\boldsymbol{x}})^T] \leq 0 \end{aligned} \tag{36}$$

where $R(\boldsymbol{x}) \equiv D(\boldsymbol{x}) D^T(\boldsymbol{x})$.

For the attitude estimation case, $\boldsymbol{\zeta}(\boldsymbol{x}) = \begin{bmatrix} 0_{1\times 3} & \boldsymbol{\beta}^T \end{bmatrix}^T$ is taken and

$$Z(\boldsymbol{x}) = \begin{bmatrix} \Xi(\boldsymbol{q}) & 0_{4\times 3} \\ 0_{3\times 3} & I_{3\times 3} \end{bmatrix} \tag{37}$$

so that

$$\boldsymbol{z} = \begin{bmatrix} -\Xi^T(\boldsymbol{q}) \hat{\boldsymbol{q}} \\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \end{bmatrix} \tag{38}$$

The first three components of $\boldsymbol{z}$ are the vector part of $\boldsymbol{q} \otimes \hat{\boldsymbol{q}}^{-1}$, the estimation error quaternion.

A significant amount of algebra using some small-error approximations shows that Eq. (36) is satisfied for

$$V(\boldsymbol{x}, \hat{\boldsymbol{x}}, t) = \gamma^2 \boldsymbol{z}^T P^{-1} \boldsymbol{z} \tag{39}$$

with the gain matrix

$$K(\hat{\boldsymbol{x}}, t) = Z(\hat{\boldsymbol{x}}) P \begin{bmatrix} H^T(\hat{\boldsymbol{q}}) \\ 0_{3\times m} \end{bmatrix} R^{-1} \tag{40}$$

where $H(\hat{\boldsymbol{q}})$ is the $m \times 3$ sensitivity matrix, which is a function of the estimated quaternion, and the 6×6 matrix P satisfies the Riccati-like equation

$$\begin{aligned} & -\dot{P} + F(\hat{\boldsymbol{x}}) P + P F^T(\hat{\boldsymbol{x}}) \\ & + P \begin{bmatrix} \gamma^{-2} I_{3\times 3} - H^T(\hat{\boldsymbol{q}}) R^{-1} H(\hat{\boldsymbol{q}}) & 0_{3\times 3} \\ 0_{3\times 3} & \gamma^{-2} I_{3\times 3} \end{bmatrix} P \\ & + \begin{bmatrix} \frac{1}{4} Q_1 & 0_{3\times 3} \\ 0_{3\times 3} & Q_2 \end{bmatrix} \leq 0 \end{aligned} \tag{41}$$

with

$$F(\hat{\boldsymbol{x}}) = -\begin{bmatrix} [(\tilde{\boldsymbol{\omega}} - \hat{\boldsymbol{\beta}})\times] & \frac{1}{2}I_{3\times 3} \\ 0_{3\times 3} & 0_{3\times 3} \end{bmatrix} \tag{42a}$$

$$Q_1 = B_1 B_1^T \quad \text{and} \quad Q_2 = B_2 B_2^T \tag{42b}$$

The scalar γ is a tuning parameter for this estimator. It can be seen that a smaller value of γ makes it more difficult to satisfy Eqs. (35) and (41), and these equations cannot be satisfied at all if γ is chosen too small. As γ becomes infinitely large, on the other hand, the equality limit of Eq. (41) goes to the usual Riccati equation of the Kalman filter, and the deterministic estimator becomes the continuous measurement limit of the MEKF [6].

IV. Methods Based on Solutions to Wahba's Problem

Filter QUEST and its variants are based on Wahba's problem [94], which is the problem of finding the proper orthogonal matrix A that minimizes the loss function

$$J(A) = \frac{1}{2}\sum_{i=1}^{m} a_i \|\tilde{\boldsymbol{b}}_i - A\boldsymbol{r}_i\|^2 \tag{43}$$

where a_i are non-negative weights and m is the total number of measurements. Writing the loss function as

$$J(A) = \lambda_0 - \text{trace}(AB^T) \tag{44}$$

with

$$\lambda_0 \equiv \sum_{i=1}^{m} a_i \tag{45}$$

and

$$B \equiv \sum_{i=1}^{m} a_i \tilde{\boldsymbol{b}}_i \boldsymbol{r}_i^T \tag{46}$$

makes it clear that $J(A)$ is minimized when $\text{trace}(AB^T)$ is maximized. This is equivalent to the orthogonal Procrustes problem [95] of finding the orthogonal matrix A that is closest to B in the Frobenius (or Euclidean, or Schur, or Hilbert–Schmidt) norm, with the proviso that A have the determinant $+1$. Shuster noted that the nine components of the attitude profile matrix B contain full information about the three attitude degrees of freedom and the six independent components of the angular error covariance matrix, and that choosing the weights to be inverse variances, $a_i = \sigma_i^{-2}$ makes Wahba's problem a maximum likelihood estimation problem [77].

There are many algorithms for solving this problem [96], of which the most useful are Davenport's q method [97] and QUEST [24]. Davenport parameterized the attitude matrix by a unit quaternion, as shown by Eq. (4), giving

$$\text{trace}\,(AB^T) = \boldsymbol{q}^T K \boldsymbol{q} \tag{47}$$

where K is the symmetric traceless matrix

$$K \equiv \begin{bmatrix} B + B^T - \text{trace}(B) I_{3\times 3} & \sum_{i=1}^{m} a_i \tilde{\boldsymbol{b}}_i \times \boldsymbol{r}_i \\ \left(\sum_{i=1}^{m} a_i \tilde{\boldsymbol{b}}_i \times \boldsymbol{r}_i\right)^T & \text{trace}(B) \end{bmatrix} = -\sum_{i=1}^{m} a_i \Omega(\tilde{\boldsymbol{b}}_i)\Gamma(\boldsymbol{r}_i) \tag{48}$$

The relation on the right-hand side of Eq. (48) can be derived by using the matrix identities in Eqs. (7) and (9). Note that the matrices Ω and Γ commute, which has some useful consequences [98]. The optimal attitude is represented by the quaternion maximizing the right-hand side of Eq. (47), subject to the unit constraint $\|\boldsymbol{q}\| = 1$. It is not difficult to see that the optimal quaternion is equal to the normalized eigenvector of K with the largest eigenvalue, that is, the solution of

$$K\boldsymbol{q}_{\text{opt}} = \lambda_{\max}\boldsymbol{q}_{\text{opt}} \tag{49}$$

With Eqs. (44) and (47), this gives the optimized loss function as

$$J(A_{\text{opt}}) = \lambda_0 - \lambda_{\max} \tag{50}$$

The QUEST algorithm is based on Shuster's observation that $\lambda_{\max}$ can be obtained by a Newton–Raphson iteration starting from λ_0 as the initial estimate, because Eq. (50) shows that $\lambda_{\max}$ is very close to λ_0 if the optimized loss function is small. In fact, a single iteration is generally sufficient. QUEST is less robust than Davenport's q method in principle, but has proved itself to be reliable in practical applications.

A. Filter Quaternion Estimator

Because a filtering algorithm is usually preferred when observations are obtained over a range of times, Shuster proposed the filter QUEST algorithm [25], based on propagating and updating B:

$$B(t_k) = \mu\Phi_{3\times 3}(t_k, t_{k-1})B(t_{k-1}) + \sum_{i=1}^{m_k} a_i \tilde{\boldsymbol{b}}_i \boldsymbol{r}_i^T \tag{51}$$

where $\Phi_{3\times 3}(t_k, t_{k-1})$ is the state transition matrix for the attitude matrix, $\mu < 1$ is a fading memory factor, and m_k is the number of observations at time t_k. The optimal attitude at time t_k is found from $B(t_k)$ by the QUEST algorithm. Shuster also formulated a smoother on the same basis.

An alternative sequential algorithm, recursive quaternion estimator (REQUEST) [26], propagates and updates Davenport's K matrix by

$$K(t_k) = \mu\Phi_{4\times 4}(t_k, t_{k-1})K(t_{k-1})\Phi_{4\times 4}^T(t_k, t_{k-1}) + \sum_{i=1}^{m_k} a_i K_i \tag{52}$$

where $\Phi_{4\times 4}(t_k, t_{k-1})$ is the quaternion state transition matrix and K_i is the Davenport matrix for a single observation:

$$K_i = \begin{bmatrix} \tilde{\boldsymbol{b}}_i \boldsymbol{r}_i^T + \boldsymbol{r}_i \tilde{\boldsymbol{b}}_i^T - \left(\tilde{\boldsymbol{b}}_i^T \boldsymbol{r}_i\right) I_{3\times 3} & (\tilde{\boldsymbol{b}}_i \times \boldsymbol{r}_i) \\ (\tilde{\boldsymbol{b}}_i \times \boldsymbol{r}_i)^T & \tilde{\boldsymbol{b}}_i^T \boldsymbol{r}_i \end{bmatrix} \tag{53}$$

Filter QUEST and REQUEST are mathematically equivalent, but filter QUEST requires fewer computations. Neither has been competitive with an EKF in practice, largely due to the suboptimality of the fading memory approximation to the effect of process noise. Computing the fading memory factor by a Kalman-gain-like algorithm gives better performance, but sacrifices much of the attractive simplicity of this method [99].

B. Extended Quaternion Estimator

Extended QUEST [27] is an algorithm that solves for the attitude along with additional parameters. This is accomplished by finding the attitude quaternion $\boldsymbol{q}_k$ and the vector of auxiliary filter states $\boldsymbol{x}_k$, along with $\boldsymbol{q}_{k-1}$, $\boldsymbol{x}_{k-1}$, and the process noise vector $\boldsymbol{w}_{k-1}$ that minimize the loss function

$$\begin{aligned} J = {} & \frac{1}{2}\sum_{i=1}^{m_k} \sigma_i^{-2} \|\tilde{\boldsymbol{b}}_i - A(\boldsymbol{q}_k)\boldsymbol{r}_i\|^2 + \frac{1}{2}\|R_{\boldsymbol{ww}(k-1)}\boldsymbol{w}_{k-1}\|^2 \\ & + \frac{1}{2}\|R_{\boldsymbol{qq}(k-1)}(\boldsymbol{q}_{k-1} - \hat{\boldsymbol{q}}_{k-1})\|^2 + \frac{1}{2}\|R_{\boldsymbol{xq}(k-1)}(\boldsymbol{q}_{k-1} - \hat{\boldsymbol{q}}_{k-1}) \\ & + R_{\boldsymbol{xx}(k-1)}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})\|^2 \end{aligned} \tag{54}$$

subject to the attitude dynamics equation,

$$\boldsymbol{q}_k = \Phi(t_k, t_{k-1}; \boldsymbol{x}_{k-1}, \boldsymbol{w}_{k-1})\boldsymbol{q}_{k-1} \tag{55}$$

the transition equation for the auxiliary filter states,

$$\boldsymbol{x}_k = \boldsymbol{f}_{\boldsymbol{x}}(t_k, t_{k-1}; \boldsymbol{q}_{k-1}, \boldsymbol{x}_{k-1}, \boldsymbol{w}_{k-1}) \tag{56}$$

and the norm constraint $\|\boldsymbol{q}_k\| = 1$. The vectors $\hat{\boldsymbol{q}}_{k-1}$ and $\hat{\boldsymbol{x}}_{k-1}$ are the a posteriori (or best) estimates of $\boldsymbol{q}$ and $\boldsymbol{x}$ at sample time t_{k-1}, and the various R matrices are weights in the loss function.

The minimization employs an extended square-root information filtering algorithm [100] that proceeds in two stages per sampling period. The first phase dynamically propagates the a posteriori estimates at stage $k-1$ to compute a priori estimates at stage k. The propagated state estimates use the full nonlinear propagation and the mean value, zero, of the process noise, as in the EKF:

$$\tilde{\boldsymbol{q}}_k = \Phi(t_k, t_{k-1}; \hat{\boldsymbol{x}}_{k-1}, \boldsymbol{0})\hat{\boldsymbol{q}}_{k-1} \tag{57}$$

and

$$\tilde{\boldsymbol{x}}_k = \boldsymbol{f}_{\boldsymbol{x}}(t_k, t_{k-1}; \hat{\boldsymbol{q}}_{k-1}, \hat{\boldsymbol{x}}_{k-1}, \boldsymbol{0}) \tag{58}$$

The result of the propagation step is a modified form of the loss function

$$\begin{aligned} J(\boldsymbol{q}_k, \boldsymbol{x}_k) &= \frac{1}{2}\sum_{i=1}^{m_k} \sigma_i^{-2}\|\tilde{\boldsymbol{b}}_i - A(\boldsymbol{q}_k)\boldsymbol{r}_i\|^2 + \frac{1}{2}\|\tilde{R}_{\boldsymbol{qq}(k)}(\boldsymbol{q}_k - \tilde{\boldsymbol{q}}_k)\|^2 \\ &\quad + \frac{1}{2}\|\tilde{R}_{\boldsymbol{xq}(k)}(\boldsymbol{q}_k - \tilde{\boldsymbol{q}}_k) + \tilde{R}_{\boldsymbol{xx}(k)}(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_k)\|^2 \end{aligned} \tag{59}$$

where the $\tilde{R}$ matrices are obtained by a QR factorization in the propagation step that employs a linearization about the a priori estimates at stage $k-1$ as in an EKF.

The second phase, the measurement update, is the novel part of extended QUEST. The optimum $\boldsymbol{x}_k$ is easily given by

$$\boldsymbol{x}_k = \tilde{\boldsymbol{x}}_k - \tilde{R}^{-1}_{\boldsymbol{xx}(k)}\tilde{R}_{\boldsymbol{xq}(k)}(\boldsymbol{q}_k - \tilde{\boldsymbol{q}}_k) \tag{60}$$

Substituting this in Eq. (59) and using Eqs. (44), (46), and (53) gives

$$\begin{aligned} J(\boldsymbol{q}_k, \hat{\boldsymbol{x}}_k) &= -\boldsymbol{q}_k^T\left(\sum_{i=1}^{m_k}\sigma_i^{-2}K_i\right)\boldsymbol{q}_k \\ &\quad + \frac{1}{2}[\tilde{R}_{\boldsymbol{qq}(k)}(\boldsymbol{q}_k - \tilde{\boldsymbol{q}}_k)]^T[\tilde{R}_{\boldsymbol{qq}(k)}(\boldsymbol{q}_k - \tilde{\boldsymbol{q}}_k)] \end{aligned} \tag{61}$$

Minimizing this loss function gives the best estimate $\hat{\boldsymbol{q}}_k$. The minimization differs from Wahba's problem because of the linear terms in $\boldsymbol{q}_k$. Substituting $\hat{\boldsymbol{q}}_k$ into Eq. (60) gives

$$\hat{\boldsymbol{x}}_k = \tilde{\boldsymbol{x}}_k - \tilde{R}^{-1}_{\boldsymbol{xx}(k)}\tilde{R}_{\boldsymbol{xq}(k)}(\hat{\boldsymbol{q}}_k - \tilde{\boldsymbol{q}}_k) \tag{62}$$

The remaining step is to express the optimized loss function in the form of Eq. (54) to prepare for the next step in the recursion. The details are in [27].

V. Two-Step Attitude Estimator

The two-step optimal estimator is an alternative to the standard EKF [23]. An implementation of the two-step optimal estimation for recursive spacecraft attitude estimation is presented in [101]. The general cost function of the two-step optimal estimator is given by

$$\begin{aligned} J_k &= \frac{1}{2}\sum_{i=0}^{k-1}\Big\{[\tilde{\boldsymbol{y}}_{i+1} - \boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})]^T R^{-1}_{i+1}[\tilde{\boldsymbol{y}}_{i+1} - \boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})] \\ &\quad + \boldsymbol{w}_i^T Q_i^{-1}\boldsymbol{w}_i\Big\} + \frac{1}{2}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^T P_0^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0) \end{aligned} \tag{63}$$

subject to the dynamics equation

$$\boldsymbol{x}_{i+1} = \boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{w}_i) \quad \text{for } i = 0, 1, \ldots, k-1 \tag{64}$$

The process noise covariance is Q_i, $\tilde{\boldsymbol{y}}_{i+1}$ is the measurement with covariance R_{i+1}, $\boldsymbol{h}_{i+1}(\boldsymbol{x}_{i+1})$ is the nonlinear measurement model, and $\hat{\boldsymbol{x}}_0$ is the a priori estimate of the state with covariance P_0. The state estimate $\hat{\boldsymbol{x}}_k$ that minimizes the above cost function is the MAP estimate if the process noise, the measurement noise, and the initial estimate error satisfy the Gaussian assumption.

The main point of the two-step optimal estimator is to define a first-step state $\mathcal{Y} = \mathcal{F}(\boldsymbol{x})$ in which the nonlinear measurement model is linear, that is, $\tilde{\boldsymbol{y}} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{v} = \mathcal{H}\mathcal{Y} + \boldsymbol{v}$, where $\boldsymbol{v}$ is the measurement noise, so that a linear measurement update of the first-step state can be applied. The first-step state $\mathcal{Y}$ is related to the desired or second-step state $\boldsymbol{x}$ and the measurement $\tilde{\boldsymbol{y}}$ through the nonlinear mapping $\mathcal{F}$ and the linear measurement matrix $\mathcal{H}$, respectively. All the information contained in the initial guess and the measurements is fused in the first-step state using a Kalman filter that performs a nonlinear and sometimes higher-order time update and a linear measurement update. The initial mean $\hat{\mathcal{Y}}_0$ and covariance $\mathcal{P}_0$ of the first-step state can be obtained from $\hat{\boldsymbol{x}}_0$ and P_0, for example, using a Monte Carlo approach. In general, $(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^T P_0^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)$ is not equivalent to $(\mathcal{Y}_0 - \hat{\mathcal{Y}}_0)^T\mathcal{P}_0^{-1}(\mathcal{Y}_0 - \hat{\mathcal{Y}}_0)$ because of the nonlinear relationship between $\boldsymbol{x}$ and $\mathcal{Y}$. The desired state estimate $\hat{\boldsymbol{x}}_k$ is obtained from the first-step state estimate $\hat{\mathcal{Y}}_k$ and its error covariance matrix $\mathcal{P}_k$ as the minimum of the cost function

$$J_{\boldsymbol{x}} = \tfrac{1}{2}[\hat{\mathcal{Y}}_k - \mathcal{F}(\boldsymbol{x})]^T\mathcal{P}_k^{-1}[\hat{\mathcal{Y}}_k - \mathcal{F}(\boldsymbol{x})] \tag{65}$$

The constraints in $\hat{\boldsymbol{x}}_k$ may also be included by use of the Lagrangian multiplier method. A numerical least-squares algorithm such as the Gauss–Newton or Levenberg–Marquardt method is used to solve the second-step minimization problem. Note that to guarantee the uniqueness of the solution, the size of the first-step state should in general be larger than or equal to that of the desired state and the covariance matrix of the first-step state should be positive definite. When there are no dynamics, that is, $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$, or initial guess, Eq. (63) reduces to

$$J_k = \frac{1}{2}\sum_{i=1}^{k}[\tilde{\boldsymbol{y}}_i - \boldsymbol{h}_i(\boldsymbol{x})]^T R_i^{-1}[\tilde{\boldsymbol{y}}_i - \boldsymbol{h}_i(\boldsymbol{x})] \tag{66}$$

The two-step attitude estimator for this special case amounts to 1) formulating the cost function in terms of $\tilde{\boldsymbol{y}}_i$ and R_i in an equivalent form in terms of $\hat{\mathcal{Y}}_k$ and $\mathcal{P}_k$ that are obtained with a linear least-squares scheme and 2) obtaining $\hat{\boldsymbol{x}}_k$ by solving the converted cost function. The first-step estimate $\hat{\mathcal{Y}}_k$ and the covariance matrix $\mathcal{P}_k$ are simply an equivalent representation of all the measurements.

In the two-step attitude estimator, the desired or second-step state is the attitude quaternion (for attitude-only estimation) or the attitude quaternion and the gyro biases (for attitude and gyro bias estimation). The assumed attitude measurement model is the unit-vector model, which is independent of the gyro biases, linear in the attitude matrix, and quadratic in the attitude quaternion. With this measurement model, the first-step state for attitude-only estimation can be chosen as

$$\mathcal{Y} = [A_{11}, A_{12}, A_{13}, A_{21}, \ldots, A_{33}]^T \tag{67}$$

where A_{ij}, $i, j = 1, 2, 3$ are the elements of the attitude matrix. The measurement matrix $\mathcal{H}$ that relates $\mathcal{Y}$ in Eq. (67) to a unit-vector observation is given by

$$\mathcal{H} = \begin{bmatrix} \boldsymbol{r}^T & \boldsymbol{0}_{1\times 3} & \boldsymbol{0}_{1\times 3} \\ \boldsymbol{0}_{1\times 3} & \boldsymbol{r}^T & \boldsymbol{0}_{1\times 3} \\ \boldsymbol{0}_{1\times 3} & \boldsymbol{0}_{1\times 3} & \boldsymbol{r}^T \end{bmatrix} \tag{68}$$

where $\boldsymbol{r}$ is the representation of the unit vector in the reference frame. The first-step state is expanded to include the gyro biases when both the attitude and the gyro biases are to be estimated.

The attitude part of the augmented first-step estimate may be understood as the usual arithmetic mean of the attitude matrix in the space of 3×3 matrices, which is not an orthogonal matrix in general. As Kasdin and Weaver noted, this first-step estimate is not to be used as an estimate of attitude because it does not take the six constraints on the attitude matrix into account [101]. The purpose of the first step is to optimally filter the vector measurement sensor noise. It is not until the second-step minimization that the best attitude estimate (as a unit quaternion) becomes available. The details of the second-step minimization of the two-step attitude estimator are in [101].

The dynamic propagation equation of the attitude part of the first-step state, $\mathcal{Y}$ in Eq. (67), is given by

$$\dot{\mathcal{Y}}(t) = \tilde{\Omega}(\boldsymbol{\omega})\mathcal{Y}(t) \tag{69}$$

where $\tilde{\Omega}(\boldsymbol{\omega})$ is given by

$$\tilde{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} 0_{3\times 3} & \omega_3 I_{3\times 3} & -\omega_2 I_{3\times 3} \\ -\omega_3 I_{3\times 3} & 0_{3\times 3} & \omega_1 I_{3\times 3} \\ \omega_2 I_{3\times 3} & -\omega_1 I_{3\times 3} & 0_{3\times 3} \end{bmatrix} \tag{70}$$

Because the attitude quaternion does not appear in the above equation, the corresponding time update of the first-step estimate is not a function of the unit quaternion estimate. Therefore, there is no need to find the unit quaternion estimate to process the filter. Only when the attitude estimate is desired does the second-step minimization need to be computed. In most of the two-step optimal estimators, however, the dynamics for the first-step state is explicitly dependent on the second-step state in a nonlinear fashion. Consequently, the time update of the first-step state is complex and computationally expensive [101].

When the angular rate $\boldsymbol{\omega}$ in Eq. (69) is perfectly known, only a linear Kalman filter needs to be implemented for the first-step state, that is, $\mathcal{Y}$ in Eq. (67). When the angular rate is obtained from noisy gyro measurements and both the attitude and the gyro biases are to be estimated, the dynamic propagation equation of the augmented first-step state becomes nonlinear because of the state-dependent process noise and the product terms of the attitude matrix elements and the gyro biases. In this case a nonlinear time-update scheme must be implemented for the first-step state.

The two-step attitude estimator for attitude-only estimation consists of:

1) Time update of the first-step estimate as a nonorthogonal matrix.

2) Measurement update of the first-step estimate as a nonorthogonal matrix.

3) Second-step minimization for the unit quaternion or the orthogonal attitude matrix (run on demand).

It has similar properties with the recursive attitude estimation algorithms in [91], although the gyro measurement and vector measurement models used therein are slightly different. In [91], the algorithm without orthogonalization corresponds to the first two steps only; the algorithm with orthogonalization corresponds to the first two steps followed by an iterative orthogonalization procedure, which is run at every measurement update. The convergence of the two algorithms in [91] is not always assured when they are initialized with the identity matrix.

VI. Unscented Filtering

The Unscented filter [28,29] works on the premise that with a fixed number of parameters it should be easier to approximate a Gaussian distribution than to approximate an arbitrary nonlinear function. The filter is derived for discrete-time nonlinear equations, where the system model is given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, k) + \boldsymbol{w}_k \tag{71a}$$

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{h}(\boldsymbol{x}_k, k) + \boldsymbol{v}_k \tag{71b}$$

Note that a continuous-time model can always be written using Eq. (71a) through an appropriate numerical integration scheme. It is again assumed that $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are zero-mean Gaussian noise processes with covariances given by Q_k and R_k, respectively.

In this section the superscript $+$ denotes an updated value and the superscript $-$ denotes a propagated value. Assuming no process noise, the formulation for the prediction equations and computation of the gain matrix is given by computing the following *sigma points*:

$$\boldsymbol{\sigma}_k \leftarrow 2n \text{ columns from } \pm\gamma\sqrt{P_k^+} \tag{72a}$$

$$\boldsymbol{\chi}_k(0) = \hat{\boldsymbol{x}}_k^+ \tag{72b}$$

$$\boldsymbol{\chi}_k(i) = \boldsymbol{\sigma}_k(i) + \hat{\boldsymbol{x}}_k^+ \tag{72c}$$

where γ is a design parameter and $\sqrt{P_k^+}$ denotes the matrix square root of P_k^+ [95]. Because of the symmetric nature of this set, its odd central moments are zero, and so its first three moments are the same as the original Gaussian distribution. The transformed set of sigma points are evaluated for each of the points by

$$\boldsymbol{\chi}_{k+1}(i) = \boldsymbol{f}[\boldsymbol{\chi}_k(i), k] \quad \text{for } i = 0, 1, \ldots, 2n \tag{73}$$

The predicted mean at time t_{k+1} for the state estimate is calculated using a weighted sum of the points $\boldsymbol{\chi}_{k+1}(i)$, which is given by

$$\hat{\boldsymbol{x}}_{k+1}^- = \sum_{i=0}^{2n} W_i^{\text{mean}} \boldsymbol{\chi}_{k+1}(i) \tag{74}$$

where W_i^{mean} is a weighting parameter. The innovations process is $\boldsymbol{\upsilon}_k \equiv \tilde{\boldsymbol{y}}_k - \hat{\boldsymbol{y}}_k^-$, where $\hat{\boldsymbol{y}}_k^- \equiv \boldsymbol{h}(\hat{\boldsymbol{x}}_k^-, k)$. The predicted covariance, output covariance, and cross correlation between $\hat{\boldsymbol{x}}_k^-$ and $\hat{\boldsymbol{y}}_k^-$ are computed by

$$P_{k+1}^- = \sum_{i=0}^{2n} W_i^{\text{cov}} \Big[\boldsymbol{\chi}_{k+1}(i) - \hat{\boldsymbol{x}}_{k+1}^-\Big]\Big[\boldsymbol{\chi}_{k+1}(i) - \hat{\boldsymbol{x}}_{k+1}^-\Big]^T \tag{75a}$$

$$P_{k+1}^{yy} = \sum_{i=0}^{2n} W_i^{\text{cov}} \Big[\boldsymbol{\gamma}_{k+1}(i) - \hat{\boldsymbol{y}}_{k+1}^-\Big]\Big[\boldsymbol{\gamma}_{k+1}(i) - \hat{\boldsymbol{y}}_{k+1}^-\Big]^T \tag{75b}$$

$$P_{k+1}^{xy} = \sum_{i=0}^{2n} W_i^{\text{cov}} \Big[\boldsymbol{\chi}_{k+1}(i) - \hat{\boldsymbol{x}}_{k+1}^-\Big]\Big[\boldsymbol{\gamma}_{k+1}(i) - \hat{\boldsymbol{y}}_{k+1}^-\Big]^T \tag{75c}$$

where W_i^{cov} is another weighting parameter and

$$\boldsymbol{\gamma}_{k+1}(i) = \boldsymbol{h}(\boldsymbol{\chi}_{k+1}(i), k+1) \tag{76}$$

Because the measurement noise appears linearly in Eq. (71b), the covariance of the innovations process at time t_{k+1} is $P_{k+1}^{\upsilon\upsilon} = P_{k+1}^{yy} + R_{k+1}$. The Kalman gain and updated covariance are rewritten in the form given by [102]

$$K_{k+1} = P_{k+1}^{xy}\Big(P_{k+1}^{\upsilon\upsilon}\Big)^{-1} \tag{77a}$$

$$P_{k+1}^+ = P_{k+1}^- - K_{k+1} P_{k+1}^{\upsilon\upsilon} K_{k+1}^T \tag{77b}$$

The state update is given by the usual EKF form:

$$\hat{\boldsymbol{x}}_{k+1}^+ = \hat{\boldsymbol{x}}_{k+1}^- + K_{k+1}\boldsymbol{\upsilon}_{k+1} \tag{78}$$

Methods to handle process noise for the computation of the predicted covariance and to handle nonlinearly appearing measurement noise are discussed in [29].

As with the standard EKF, using the UF directly with a quaternion parameterization of the attitude yields a nonunit quaternion estimate, as seen by Eq. (74). To overcome this problem an unconstrained three-component vector is used, based on the generalized Rodrigues parameters [81] (GRPs) to represent an attitude error quaternion [31]. The algorithm is called the unscented quaternion estimator, or USQUE. The state vector includes the attitude error and bias vectors:

$$\boldsymbol{\chi}_k(0) = \hat{\boldsymbol{x}}_k^+ \equiv \begin{bmatrix} \boldsymbol{\delta}\hat{\boldsymbol{p}}_k^+ \\ \hat{\boldsymbol{\beta}}_k^+ \end{bmatrix} \tag{79}$$

where the estimated error-GRP $\boldsymbol{\delta}\hat{\boldsymbol{p}}_k$ is used to propagate and update a nominal quaternion. This estimate, as well as the corresponding sigma points, can be used to form error quaternions, denoted by $\boldsymbol{\delta q}_k^+(i)$, through a simple transformation from GRPs to quaternions. Then, the following quaternions are computed:

$$\hat{\boldsymbol{q}}_k^+(0) = \hat{\boldsymbol{q}}_k^+ \tag{80a}$$

$$\hat{\boldsymbol{q}}_k^+(i) = \boldsymbol{\delta q}_k^+(i) \otimes \hat{\boldsymbol{q}}_k^+, \qquad i = 1, 2, \ldots, 12 \tag{80b}$$

A reset of the attitude error to zero after the previous update is required, which is used to move information from one part of the estimate to another part [7]. This reset rotates the reference frame for the covariance, and so it is expected the covariance be rotated, even though no new information is added. But the covariance depends on the assumed statistics of the measurements, not on the actual measurements. Therefore, because the update is zero mean, the mean rotation caused by the reset is actually zero, and so the covariance is in fact not affected by the reset. The quaternions in Eq. (80) are propagated using Eq. (7) with the estimated angular rate. The propagated error quaternions are then computed using

$$\boldsymbol{\delta q}_{k+1}^-(i) = \hat{\boldsymbol{q}}_{k+1}^-(i) \otimes \left[\hat{\boldsymbol{q}}_{k+1}^-(0)\right]^{-1}, \qquad i = 0, 1, \ldots, 12 \tag{81}$$

Note that $\boldsymbol{\delta q}_{k+1}^-(0)$ is the identity quaternion. Finally, the propagated error-GRPs are computed using a simple transformation from quaternions to GRPs. The error-GRPs can be propagated directly, however, this approach requires the integration of nonlinear equations. The advantage of converting the GRPs to quaternions is that a closed-form discrete-time solution exists for Eq. (7).

VII. Particle Filters

There is no such thing as "*the* PF," just as there is no such thing as "*the* EKF" [103]. Particle filters comprise a very broad class of suboptimal nonlinear filters based on sequential Monte Carlo simulations, in which the distributions are approximated by weighted particles (random samples) that are generated using pseudorandom number generators. The computational expense and attainable estimation accuracy of PFs vary greatly. Numerous theories, improving strategies, and applications of the PFs can be found in [32,33,104].

The general discrete-time model used in PFs is given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}_k(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{w}_k) \tag{82a}$$

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{v}_k) \tag{82b}$$

The main difference between PFs and other filters is the process noise $\boldsymbol{w}_k$ and the measurement noise $\boldsymbol{v}_k$ are not necessarily assumed to be normally distributed processes. The distributions of $\boldsymbol{x}_0$, $\boldsymbol{w}_k$, and $\boldsymbol{v}_k$, denoted by $p(\boldsymbol{x}_0)$, $p(\boldsymbol{w}_k)$, and $p(\boldsymbol{v}_k)$, respectively, are assumed to be known and mutually independent. The probabilities $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$ and $p(\tilde{\boldsymbol{y}}_k|\boldsymbol{x}_k)$ can be derived from the above model and are assumed to be available for sampling and evaluation.

The variables $\boldsymbol{X}_k$ and $\tilde{\boldsymbol{Y}}_k$ are used to denote the state trajectory $\{\boldsymbol{x}_j\}_{j=0}^k$ and measurement history $\{\tilde{\boldsymbol{y}}_j\}_{j=1}^k$, respectively. From the sampling perspective, the empirical, discrete approximation of the posterior distribution $p(\boldsymbol{x}_k|\tilde{\boldsymbol{Y}}_k)$ with N weighted particles $\{\boldsymbol{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ is given by [32]

$$p_N(d\boldsymbol{x}_k|\tilde{\boldsymbol{Y}}_k) \approx \sum_{i=1}^N w_k^{(i)} \delta_{\boldsymbol{x}_k^{(i)}}(d\boldsymbol{x}_k) \tag{83}$$

where $\boldsymbol{x}_k^{(i)}$ are the particles drawn from the importance function or proposal distribution $q(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k^{(i)}, \tilde{\boldsymbol{Y}}_{k+1})$, $w_k^{(i)}$ are the normalized importance weights, satisfying $\sum_{i=1}^N w_k^{(i)} = 1$, and $\delta_{\boldsymbol{x}_k^{(i)}}(d\boldsymbol{x}_k)$ denotes the delta-Dirac mass located in $\boldsymbol{x}_k^{(i)}$. The importance function can be chosen from a large class of distributions. It is only required that the support of the importance function include the support of $p(\boldsymbol{x}_k|\tilde{\boldsymbol{Y}}_k)$. The importance weight is the ratio of the posterior distribution to the importance function evaluated at $\boldsymbol{x}_k^{(i)}$. The expectation of a known function $\boldsymbol{f}(\boldsymbol{x}_k)$ with respect to $p(\boldsymbol{x}_k|\tilde{\boldsymbol{Y}}_k)$ is then approximated by $\sum_{i=1}^N w_k^{(i)} \boldsymbol{f}(\boldsymbol{x}_k^{(i)})$ [32]. For example, the approximation to the arithmetic mean of $\boldsymbol{x}_k$ is $\sum_{i=1}^N w^{(i)}\boldsymbol{x}^{(i)}$. Crisan and Doucet showed that the upper bound on the variance of the estimation error of the expectation has the form $c\mathcal{O}(N^{-1})$, with c a constant [104,105]. Daum argued that c in the upper bound depends heavily on the state vector dimension [103,104].

A PF updates the particle representation $\{\boldsymbol{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ in a recursive manner. A cycle of a generic PF includes [104]

Sequential Importance Sampling: 1) For $i = 1, \ldots, N$, sample $\boldsymbol{x}_{k+1}^{(i)}$ from the importance function $q(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k^{(i)}, \tilde{\boldsymbol{Y}}_{k+1})$; 2) For $i = 1, \ldots, N$, evaluate and normalize the importance weights

$$w_{k+1}^{(i)} \propto w_k^{(i)} \frac{p\left(\tilde{\boldsymbol{y}}_{k+1}|\boldsymbol{x}_{k+1}^{(i)}\right) p\left(\boldsymbol{x}_{k+1}^{(i)}|\boldsymbol{x}_k^{(i)}\right)}{q\left(\boldsymbol{x}_{k+1}^{(i)}|\boldsymbol{X}_k^{(i)}, \tilde{\boldsymbol{Y}}_{k+1}\right)} \tag{84}$$

Resampling: Multiply/discard particles $\{\boldsymbol{x}_{k+1}^{(i)}\}_{i=1}^N$ with respect to high/low importance weights $w_{k+1}^{(i)}$ to obtain N new particles $\{\boldsymbol{x}_{k+1}^{(i)}\}_{i=1}^N$ with equal weights.

A popular suboptimal choice of the importance function is $q(\boldsymbol{x}_{k+1}|\boldsymbol{X}_k^{(i)}, \tilde{\boldsymbol{Y}}_{k+1}) = p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{(i)})$. The PF with this importance function is known as the bootstrap filter (BF). Sampling $\boldsymbol{x}_{k+1}^{(i)}$ from $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{(i)})$ is equivalent to the dynamic propagation of $\boldsymbol{x}_k^{(i)}$ to time t_{k+1}. The information contained in the measurement $\tilde{\boldsymbol{y}}_{k+1}$ is not employed in the sampling process. The corresponding update of the importance weight, Eq. (84), has a simple form, that is, $w_{k+1}^{(i)} \propto w_k^{(i)} p(\tilde{\boldsymbol{y}}_{k+1}|\boldsymbol{x}_{k+1}^{(i)})$, which is appealing especially when the evaluation of $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$ is difficult. Note that this choice becomes inefficient when the overlap between $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{(i)})$ and $p(\tilde{\boldsymbol{y}}_{k+1}|\boldsymbol{x}_{k+1})$ is small. In contrast with this simple importance function, the optimal importance function $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{(i)}, \tilde{\boldsymbol{y}}_{k+1})$ that minimizes the variance of the importance weight $w_k^{(i)}$ conditional upon $\boldsymbol{x}_k^{(i)}$ and $\tilde{\boldsymbol{y}}_{k+1}$ fully incorporates the latest measurement $\tilde{\boldsymbol{y}}_{k+1}$, but usually cannot be evaluated exactly or have samples drawn from it. The resampling step alleviates the inherent particle degeneracy of sequential importance sampling, but also reduces the number of distinct particles, which is often called the problem of particle impoverishment. Simple remedies for the impoverishment problem include roughening and regularization [104].

Particle filters are superior to conventional nonlinear filters for strongly nonlinear and non-Gaussian filtering problems, but they are also known for being computationally expensive. A main issue of PFs is the computational complexity for high-dimensional systems. For a 100-dimensional linear and Gaussian system, a linear Kalman filter can give the optimal Bayesian estimate, but the simulation-based PF without employing the linear Gaussian structure of the system may be unable to produce useful results.

Recent applications of PFs in spacecraft attitude estimation are reported in [34,35,106,107]. The PFs employed therein are essentially the BF, in which the attitude is propagated through attitude kinematics or dynamics and the measurements are only used to update the importance weights. In [34], a simple BF was designed to simultaneously estimate the attitude and the gyro biases (or the attitude and the attitude rate in gyroless applications). The uniform attitude distribution is used as the initial attitude distribution for the case of no initial attitude knowledge. A gradually decreasing measurement variance is used in the computation of the importance weights. This application showed that a simple BF with careful design can tackle a nontrivial six-dimensional attitude estimation problem. In [35,106], the genetic algorithm-embedded quaternion particle filter (GA-QPF) is presented for attitude and gyro bias estimation. The adaptive version of the GA-QPF is given in [107], which estimates the measurement noise distribution on the fly, along with the estimation of the spacecraft attitude and gyro biases. The noise distribution estimation scheme is based on an analysis of the filter-generated innovations process.

In [35,106,107] the problem of attitude and gyro bias estimation is divided into two easier problems. The resulting GA-QPF has an interlaced structure. The quaternion particle filter (QPF) that estimates the attitude quaternion for a given gyro bias estimate is interlaced with an external maximum likelihood estimator for the estimation of the gyro bias. The likelihood function for gyro bias estimation is approximated as [35]

$$\prod_{j=k_1}^{k_2} p[\tilde{\boldsymbol{y}}_j | \Phi_{4\times 4}(\tilde{\boldsymbol{\omega}}_j - \boldsymbol{\beta})\hat{\boldsymbol{q}}_{j-1}] \tag{85}$$

where $\Phi_{4\times 4}(\cdot)$ is the quaternion transition matrix derived from the continuous-time quaternion kinematics and $\hat{\boldsymbol{q}}_{j-1}$ is obtained from the QPF. In the likelihood function, the gyro bias $\boldsymbol{\beta}$ is treated as constant over $[t_{k_1}, t_{k_2}]$. The dynamic gyro noise model, for example Eq. (14), is not incorporated into the likelihood function. The GA algorithm maximizes the likelihood sequentially in time. The number of attitude particles is dramatically reduced by a simple initialization procedure of the QPF. The idea is based on the fact that the first vector observation defines a quaternion of rotation up to 1 degree of freedom. This degree of freedom is used to generate the initial set of particles from the first observation only.

Unlike the filters that work directly with the state mean and covariance, PFs compute the mean and covariance as derived quantities of the particle representation $\{\boldsymbol{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$. The interactions between the mean and covariance as in the EKFs no longer exist in PFs, and the issues such as attitude representations and attitude error definitions are much less significant in PFs than in EKFs. Three methods for computing the attitude estimate have been proposed:

1) As the usual arithmetic mean of three-component attitude representations such as the MRPs [34], or

2) As the attitude particle with the largest importance weight [107], or

3) As the minimum of the cost function [107]

$$\sum_{i=1}^{N} w_k^{(i)} \left\| A\left(\boldsymbol{q}_k^{(i)}\right) - \hat{A} \right\|^2 \tag{86}$$

subject to $\hat{A}^T\hat{A} = \hat{A}\hat{A}^T = I_{3\times 3}$.

The third method is the minimum mean square error (MMSE) estimate in SO(3) and can be computed using the singular value decomposition method [107]. The metric in SO(3) associated with the cost function of the MMSE estimate is given by [108]

$$d_F(A_1, A_2) = \|A_1 - A_2\| \tag{87}$$

which is in general not identical with the Riemannian metric in SO(3) corresponding to the length of the shortest geodesic curve, given by [108]

$$d_R(A_1, A_2) = \left\| \log\left(A_1^T A_2\right) \right\| \tag{88}$$

The third method is superior to the other two but is also more computationally expensive. A problem with the second method is that the particle with the largest importance weight does not necessarily have the maximum a posteriori probability because the importance weight is defined as the ratio between the posterior distribution and the importance function. The choice of the importance function may have great impact on the maximum of the importance weights and the corresponding attitude estimate. The second method for computing the attitude estimate is also known to be inaccurate. The first method is not attitude-parameterization independent or invariant under rotations; it treats the attitude particles in a three-component attitude representation like real vectors in the Euclidian space. This approximation is good when the attitude particles are distributed over a small angle.

VIII. Orthogonal Attitude Filter

The orthogonal filter [39] represents the attitude by an orthogonal rotation matrix, rather than by some parameterization of the rotation matrix. This avoids questions about singularities of representations or covariance matrices arising in other filters, and has the additional advantage of providing a consistent initialization for a completely unknown initial attitude, owing to the fact that SO(3), the group of rotation matrices, is a compact space [109]. The PDF is a non-Gaussian function defined on the Cartesian product of SO(3) and the Euclidean space $\mathbb{R}^N$ of bias parameters. The Fokker–Planck equation [38] propagates the PDF between measurements and Bayes' formula [1] incorporates measurement information. This approach is related to earlier work by Daum [37] and Lo [110–112]. It is well known that the Fokker–Planck equation for linear dynamics and Bayes' formula for a linear measurement model lead to the usual Kalman filter with a Gaussian PDF.

The non-Gaussian PDF of the attitude on SO(3) is defined by

$$p_{A_\mu|Y_v}(A) = \exp[-J_{A_\mu|Y_v}(A)] \tag{89}$$

where the negative-log-likelihood function is given by

$$J_{A_\mu|Y_v}(A) = -\mathrm{trace}\left(B_{\mu|v}^T A\right) + \text{constant} \tag{90}$$

The real 3×3 matrix $B_{\mu|v}$ has the correct number of free parameters to represent the mean and covariance of an attitude state, as in Wahba's problem. In fact, a filter based on this PDF will look very much like filter QUEST. Equation (90) contains the first two terms, $\ell = 0$ and $\ell = 1$, of an expansion of the negative-log-likelihood function in the irreducible representations of SO(3) [113]. An attitude PDF of this form was first considered by Lo [110–112], who referred to it as an exponential Fourier density. Lo included the higher-order $\ell = 2$ term, but treated neither process noise nor nonattitude bias parameters.

Three attitude estimates can be defined for this PDF: the conditional expectation

$$\bar{A}_{\mu/v} \equiv \int_{\mathrm{SO}(3)} A p_{A_\mu|Y_v}(A)\, \mathrm{d}\mu(A) \tag{91}$$

the MAP estimate

$$\hat{A}_{\mu/v} \equiv \arg\max_{A\in\mathrm{SO}(3)}[p_{A_\mu|Y_v}(A)] \tag{92}$$

and the MMSE estimate

$$A_{\mu/v}^{\mathrm{MMSE}} \equiv \arg\min_{A'\in\mathrm{SO}(3)}\left[\int_{\mathrm{SO}(3)} \|A - A'\|^2 p_{A_\mu|Y_v}(A)\, \mathrm{d}\mu(A)\right] \tag{93}$$

where $\mathrm{d}\mu(A)$ is an invariant volume element on SO(3). The conditional expectation is not an acceptable attitude estimate, because it is not an orthogonal matrix in general. The MAP and MMSE attitude estimates are orthogonal matrices by definition, and they can be shown to be identical for this PDF [39].

A PDF describing uncorrelated bias parameters obeying Gaussian statistics and an attitude with a PDF specified by Eqs. (89) and (90) would be the exponential of the sum of the right-hand-side of Eq. (90) and a quadratic function of the bias parameters. The simplest generalization to include correlations between the bias vector and the attitude matrix in the orthogonal filter is

$$J_{A_\mu,\boldsymbol{x}_\mu|Y_v}(A, \boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^T F_{\mu|v}^x \boldsymbol{x} - \boldsymbol{\psi}_{\mu|v}^T \boldsymbol{x} - \mathrm{trace}\left[B_{\mu|v}^T(\boldsymbol{x})A\right] + \text{constant} \tag{94}$$

where

$$B_{\mu|v}(\boldsymbol{x}) \equiv B_{\mu|v,0} + \sum_{k=1}^{N} B_{\mu|v,k} x_k \tag{95}$$

where x_k is the kth component of $\boldsymbol{x}$. Propagation and update equations for the parameters $F^x_{t|\mu}$, $\boldsymbol{\psi}_{t|\mu}$, and $B_{\mu|v,k}$ for $k = 0, \ldots, N$ are derived from the Fokker–Planck equation and Bayes' formula, respectively. The MAP and MMSE estimates will differ for the general correlated problem. Therefore, the method uses the more easily computed MAP estimates, which are found by simultaneously satisfying the equations

$$\hat{A}_{\mu/v} = \arg \min_{A \in \mathrm{SO}(3)} [J_{A_\mu, \boldsymbol{x}_\mu | Y_v}(A, \hat{\boldsymbol{x}}_{\mu|v})] \tag{96}$$

and

$$\hat{\boldsymbol{x}}_{\mu/v} = \arg \min_{\boldsymbol{x} \in \mathbb{R}^N} [J_{A_\mu, \boldsymbol{x}_\mu | Y_v}(\hat{A}_{\mu|v}, \boldsymbol{x})] \tag{97}$$

Defining a vector function $\boldsymbol{\eta}_{\mu|v}(A)$ with components given by

$$[\boldsymbol{\eta}_{\mu|v}(A)]_k = \mathrm{trace}\left(B^T_{\mu|v,k} A\right) \quad \text{for } k = 1, \ldots, N \tag{98}$$

allows Eq. (97) to be rewritten as

$$\begin{aligned} \hat{\boldsymbol{x}}_{\mu/v} = \arg \min_{\boldsymbol{x} \in \mathbb{R}^N} & \Big\{ \tfrac{1}{2} \boldsymbol{x}^T F^x_{\mu|v} \boldsymbol{x} - \left[\boldsymbol{\psi}_{\mu|v} + \boldsymbol{\eta}_{\mu|v}(\hat{A}_{\mu|v}) \right]^T \boldsymbol{x} \\ & - \mathrm{trace}\left(B^T_{\mu|v,0} \hat{A}_{\mu|v} \right) \Big\} = \left(F^x_{\mu|v} \right)^{-1} \left[\boldsymbol{\psi}_{\mu|v} + \boldsymbol{\eta}_{\mu|v}(\hat{A}_{\mu|v}) \right] \end{aligned} \tag{99}$$

The attitude estimate $\hat{A}_{\mu|v}$ can be found by maximizing $\mathrm{trace}[B^T_{\mu|v}(\hat{\boldsymbol{x}})A]$ for some initial guess for $\hat{\boldsymbol{x}}_{\mu|v}$, using one of the algorithms developed for Wahba's problem, then updating $\hat{\boldsymbol{x}}_{\mu|v}$ using Eq. (99), and iterating this procedure until it converges.

Because the Fokker–Planck equation is nonlinear, the $\ell = 1$ term in the PDF leads to $\ell = 2$ terms in the propagation, which cannot be accommodated in the filter. More realistic measurement models may also require the $\ell = 2$ irreducible representation to be included in the PDF. If $\ell = 2$ terms are included in the PDF, however, the nonlinear Fokker–Planck equation will introduce $\ell = 3$ and $\ell = 4$ terms that would have to be ignored. It appears that no PDF including a finite number of irreducible representations of SO(3) can provide an exact solution of the Fokker–Planck equation, and it remains to be seen if a consistent algorithm of this type can be found.

IX. Predictive Filtering

In the nonlinear predictive filter it is assumed that the state and output estimates are given by a preliminary model and a to-be-determined model-error vector, given by [36]

$$\dot{\hat{\boldsymbol{x}}}(t) = \boldsymbol{f}[\hat{\boldsymbol{x}}(t)] + G[\hat{\boldsymbol{x}}(t)]\boldsymbol{d}(t) \tag{100a}$$

$$\hat{\boldsymbol{y}}(t) = \boldsymbol{h}[\hat{\boldsymbol{x}}(t)] \tag{100b}$$

where $\boldsymbol{d}(t)$ is the model error, which may include both model variations and external disturbances. A Taylor series expansion of the output estimate in Eq. (100b) is given by

$$\hat{\boldsymbol{y}}(t + \Delta t) = \hat{\boldsymbol{y}}(t) + \boldsymbol{z}[\hat{\boldsymbol{x}}(t), \Delta t] + \Lambda(\Delta t) S[\hat{\boldsymbol{x}}(t)] \boldsymbol{d}(t) \tag{101}$$

where the ith element of $\boldsymbol{z}[\hat{\boldsymbol{x}}(t), \Delta t]$ is given by

$$z_i[\hat{\boldsymbol{x}}(t), \Delta t] = \sum_{j=1}^{p_i} \frac{\Delta t^j}{j!} L^j_f(h_i) \tag{102}$$

where p_i, $i = 1, 2, \ldots, m$, is the lowest order of the derivative of $h_i[\hat{\boldsymbol{x}}(t)]$ in which any component of $\boldsymbol{d}(t)$ first appears due to successive differentiation and substitution for $\dot{\hat{x}}_i(t)$ on the right side, and $L^j_f(h_i)$ is a jth-order Lie derivative [114]. The matrix $\Lambda(\Delta t)$ is diagonal with elements given by $\Delta t^{p_i}/p_i!$, and the ith row of $S[\hat{\boldsymbol{x}}(t)]$ is given by

$$S_i[\hat{\boldsymbol{x}}(t)] = \left\{ L_{g_1}\left[L_f^{p_i - 1}(h_i) \right], \ldots, L_{g_q}\left[L_f^{p_i - 1}(h_i) \right] \right\} \tag{103}$$

where q is the number of columns of $G[\hat{\boldsymbol{x}}(t)]$ and $L_{g_j}[L_f^{p_i-1}(h_i)]$ is another Lie derivative. Equation (103) is in essence a generalized sensitivity matrix for nonlinear systems.

A cost functional consisting of the weighted sum square of the measurement-minus-estimate residuals plus the weighted sum square of the model correction term is minimized to determine $\boldsymbol{d}(t)$. This yields [36,115]

$$\begin{aligned} \boldsymbol{d}(t) = & (\{\Lambda(\Delta t) S[\hat{\boldsymbol{x}}(t)]\}^T R^{-1} \{\Lambda(\Delta t) S[\hat{\boldsymbol{x}}(t)]\} + W)^{-1} \\ & \times \{\Lambda(\Delta t) S[\hat{\boldsymbol{x}}(t)]\}^T R^{-1} \{\tilde{\boldsymbol{y}}(t + \Delta t) - \hat{\boldsymbol{y}}(t) - \boldsymbol{z}[\hat{\boldsymbol{x}}(t), \Delta t]\} \end{aligned} \tag{104}$$

where the matrix R is the measurement covariance and the matrix W serves to weight the amount of model error added to correct the assumed model in Eq. (100). As W decreases, more model error is added to correct the model, so that the estimates more closely follow the measurements. As W increases, less model error is added, so that the estimates more closely follow the propagated model. An optimal W can be computed using an output covariance constraint which is that the covariance of the measurement residual matches the actual measurement covariance in a statistical sense [36]. Equation (104) is used in Eq. (100a) to perform a nonlinear propagation of the state estimates to time t_k, then the measurement is processed at time t_{k+1} to find the new $\boldsymbol{d}(t)$ in $[t_k, t_{k+1}]$, and then the state estimates are propagated to time t_{k+1}.

An example of the predictive filter involves using the following model to estimate the quaternion:

$$\dot{\hat{\boldsymbol{q}}} = \frac{1}{2} \Xi(\hat{\boldsymbol{q}}) \boldsymbol{d} \tag{105}$$

where $\boldsymbol{d}$ is determined from attitude measurements, such as GPS carrier-phase difference [116]. The advantage of the predictive filter over the EKF for attitude estimation applications is that the linearization is performed at the output, not at the system dynamics. Therefore, the issue of quaternion normalization is never a problem in the predictive filter, because $\boldsymbol{d}$ is used to propagate the quaternion kinematics directly. Another advantage of the predictive filter is that it can be used to provide a point-by-point solution by setting $W = 0$. Other applications of the predictive filters for attitude estimation are shown in [117,118].

X. Nonlinear Observers

Many nonlinear observers exist that provide attitude and angular rate estimates [47,48,50–52,119]. Each has its various advantages and disadvantages. In this section the observer designed by Thienel et al. [52,119] is shown, because it is the most recent of the aforementioned references. In their approach, the measured angular rate of Eq. (14a) is rewritten as

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \boldsymbol{\beta} \tag{106}$$

where the vector $\boldsymbol{\beta}$ is now assumed to be constant. The estimated angular rate is given by $\hat{\boldsymbol{\omega}} = \tilde{\boldsymbol{\omega}} - \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the estimated bias. Noiseless observations of the true quaternion, $\boldsymbol{q}$, are assumed in [52], but this assumption is relaxed in [119]. The error quaternion between the "measured" quaternion and the estimated quaternion follows Eq. (20) with

$$\boldsymbol{\delta q} \equiv \begin{bmatrix} \boldsymbol{\delta \varrho} \\ \delta q_4 \end{bmatrix} = \boldsymbol{q} \otimes \hat{\boldsymbol{q}}^{-1} \tag{107}$$

The nonlinear observer for the quaternion and bias is given by

$$\dot{\hat{\boldsymbol{q}}} = \tfrac{1}{2} \Xi(\hat{\boldsymbol{q}}) A^T(\boldsymbol{\delta q}) [\hat{\boldsymbol{\omega}} + k \boldsymbol{\delta \varrho} \,\mathrm{sign}(\delta q_4)] \tag{108a}$$

$$\dot{\hat{\boldsymbol{\beta}}} = -\tfrac{1}{2}\boldsymbol{\delta\varrho}\,\mathrm{sign}(\delta q_4) \tag{108b}$$

where k is any positive constant. The error dynamics of the observer can be shown to be given by

$$\boldsymbol{\delta}\,\dot{\boldsymbol{q}} = -\tfrac{1}{2}\Xi(\boldsymbol{\delta q})[\Delta\boldsymbol{\beta} + k\boldsymbol{\delta\varrho}\,\mathrm{sign}(\delta q_4)] \tag{109a}$$

$$\Delta\dot{\boldsymbol{\beta}} = \tfrac{1}{2}\boldsymbol{\delta\varrho}\,\mathrm{sign}(\delta q_4) \tag{109b}$$

where $\Delta\boldsymbol{\beta} \equiv \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$. The equilibrium states for Eq. (109) are $\boldsymbol{\delta q} = [0\ 0\ 0\ \pm 1]^T$ and $\Delta\boldsymbol{\beta} = [0\ 0\ 0]^T$.

To prove global stability of the observer, the following candidate Lyapunov function is chosen:

$$V = \tfrac{1}{2}\Delta\boldsymbol{\beta}^T\Delta\boldsymbol{\beta} + \tfrac{1}{2}\begin{cases} (\delta q_4 - 1)^2 + \boldsymbol{\delta\varrho}^T\boldsymbol{\delta\varrho}, & \delta q_4 \geq 0 \\ (\delta q_4 + 1)^2 + \boldsymbol{\delta\varrho}^T\boldsymbol{\delta\varrho}, & \delta q_4 < 0 \end{cases} \tag{110}$$

Taking the time derivative of Eq. (110) and using Eq. (109) leads to

$$\dot{V} = -\tfrac{k}{2}\boldsymbol{\delta\varrho}^T\boldsymbol{\delta\varrho} \tag{111}$$

for all t. This establishes that $\Delta\boldsymbol{\beta}$, $\boldsymbol{\delta\varrho}$, and δq_4 are globally uniformly bounded. The second derivative is given by

$$\ddot{V} = \tfrac{k}{2}\boldsymbol{\delta\varrho}^T(\delta q_4 I_{3\times 3} + [\boldsymbol{\delta\varrho}\times])[\Delta\boldsymbol{\beta} + k\boldsymbol{\delta\varrho}\,\mathrm{sign}(\delta q_4)] \tag{112}$$

which is also bounded. Barbalat's lemma [120] then shows that $\|\boldsymbol{\delta\varrho}\| \to 0$ as $t \to \infty$.

Nonlinear observers are especially useful because they are often accompanied with global stability proofs. The property of guaranteed convergence from any initial condition is especially desired by designers of spacecraft attitude estimation applications. The observers in [47–52] all require an attitude measurement, which limits their use to cases where a point-by-point determined attitude is known. Although nonlinear observers are still in their infancy, these methods show great promise for future applications.

XI. Adaptive Methods

Many adaptive methods exist that either update noise covariances in a filter design [17,54–56], or update model parameters through least-squares techniques [58–61] or by using nonlinear techniques [62–64]. In this section a noise adaptive approach using a quaternion Kalman filter [17], and a nonlinear adaptive approach to determine both inertia parameters and constant disturbance [62] are shown.

A. Noise Adaptive Approach

The adaptive filter described in this section is based on a linear pseudomeasurement, given by Eq. (25). The adaptive filter processes the measurement residuals to optimally compensate for system errors. The case of process noise adaptive estimation is considered in [17]. The process noise covariance matrix is assumed to be a scalar times identity matrix $Q_k = \eta I_{3\times 3}$, where η is the to-be-estimated parameter using an adaptive scheme. The process noise covariance is initialized using $Q_k = I_{3\times 3}$. Denoting the sensitivity matrix that multiplies $\boldsymbol{q}$ in Eq. (25) as H, the m–i time-step back residual is given by

$$\boldsymbol{v}_k^{\,i} = -H_k^i\hat{\boldsymbol{q}}_k^i \tag{113}$$

for $i = 1, 2, \ldots, m$, where H_k^i is computed from a single measurement and $\hat{\boldsymbol{q}}_k^i$ is the propagated quaternion estimate. Note that Eq. (113) implies that the pseudomeasurement is zero [17]. The sample mean of m predicted residuals is defined by

$$\bar{\boldsymbol{v}}^{\,m} \equiv \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{v}_k^i \tag{114}$$

The squared residual sample mean, denoted by the matrix M^m is computed as

$$M^m = \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{v}_k^i\right)\left(\boldsymbol{v}_k^i\right)^T \tag{115}$$

and the covariance of $\bar{\boldsymbol{v}}^m$ is denoted by S^m. The adaptive procedure determines η that solves the following minimization problem:

$$\min_{\eta\geq 0}\{J(\eta) = \|M^m - S^m(\eta)\|^2\} \tag{116}$$

where the norm $\|\cdot\|$ denotes the Frobenius norm, that is, $\|M\|^2 = \mathrm{trace}(MM^T)$. The solution to this minimization problem is given through a series of steps. First, the following variables are initialized: $\hat{\boldsymbol{q}}_k^0 = \hat{\boldsymbol{q}}_k^+$, $\mathcal{P}_k^0 = P_k^+$, $M^0 = 0_{4\times 4}$, $L^0 = 0_{4\times 4}$, and $\mathcal{M}^0 = 0_{4\times 4}$, where $\hat{\boldsymbol{q}}_k^+$ and P_k^+ are the updated quaternion estimate and covariance, respectively. Then, the following steps are given, starting with $i = 1$:

1) Compute $\Xi(\hat{\boldsymbol{q}}_k^{i-1})$, where Ξ is defined by Eq. (5a).

2) Propagate the quaternion and covariance using

$$\hat{\boldsymbol{q}}_k^{\,i} = \Phi_k^{i-1}\hat{\boldsymbol{q}}_k^{i-1} \tag{117a}$$

$$\mathcal{P}_k^{\,i} = \left(\Phi_k^{i-1}\right)\mathcal{P}_k^{i-1}\left(\Phi_k^{i-1}\right)^T \tag{117b}$$

where Φ_k^{i-1} is the state transition matrix from the quaternion kinematics.

3) Compute $\Xi(\hat{\boldsymbol{q}}_k^i)$ and $\boldsymbol{v}_k^i$ from Eq. (113), along with the following variables:

$$M^i = \frac{i-1}{i}M^{i-1} + \frac{1}{i}\left(\boldsymbol{v}_k^i\right)\left(\boldsymbol{v}_k^i\right)^T \tag{118a}$$

$$\mathcal{R}_k^{\,i} = \tfrac{1}{4}\left[\Xi\left(\hat{\boldsymbol{q}}_k^i\right)\right]R_k^i\left[\Xi\left(\hat{\boldsymbol{q}}_k^i\right)\right]^T + \alpha I_{4\times 4} \tag{118b}$$

$$\begin{aligned}\mathcal{Z}^i = \Big\{&\left(\Phi_k^{i-1}\right)\left[\Xi\left(\hat{\boldsymbol{q}}_k^{i-1}\right)\right]\left[\Xi\left(\hat{\boldsymbol{q}}_k^{i-1}\right)\right]^T\left(\Phi_k^{i-1}\right)^T \\ &+ \left[\Xi\left(\hat{\boldsymbol{q}}_k^i\right)\right]\left[\Xi\left(\hat{\boldsymbol{q}}_k^i\right)\right]^T\Big\}\frac{\Delta t^2}{4}\end{aligned} \tag{118c}$$

$$\mathcal{M}^{\,i} = \mathcal{M}^{i-1} + \left(H_k^i\right)\mathcal{P}_k^i\left(H_k^i\right)^T + \mathcal{R}_k^i \tag{118d}$$

$$L^i = L^{i-1} + \left(H_k^i\right)\mathcal{Z}_i\left(H_k^i\right)^T \tag{118e}$$

where R_k^i is the measurement covariance and α is a small number to ensure that $\mathcal{R}_k^i$ is nonsingular. The procedure continues through $i \leq m$. The estimate for η, denoted by $\hat{\eta}$, is computed by

$$\hat{\eta} = \frac{\mathrm{trace}[(M^m - \mathcal{M}^m)(L^m)^T]}{\mathrm{trace}[(L^m)(L^m)^T]} \tag{119}$$

Then, the process noise covariance is updated. This covariance is then converted into a 4×4 matrix using the matrix Ξ, which is then used in the additive filter after the normalization stage.

B. Nonlinear Adaptive Approach

The nonlinear adaptive approach described in this section estimates both the inertia matrix components and a constant vector of unknown disturbances [62]. The kinematics are given by the vector of MRPs [66], denoted by $\boldsymbol{p}$, and Euler's dynamics equations [121]:

$$\dot{\boldsymbol{p}} = \tfrac{1}{4}\{(1 - \boldsymbol{p}^T\boldsymbol{p})I_{3\times 3} + 2[\boldsymbol{p}\times] + 2\boldsymbol{p}\boldsymbol{p}^T\}\boldsymbol{\omega} \equiv \tfrac{1}{4}B(\boldsymbol{p})\boldsymbol{\omega} \tag{120a}$$

$$\mathcal{J}\dot{\boldsymbol{\omega}} + [\boldsymbol{\omega}\times]\mathcal{J}\boldsymbol{\omega} = \boldsymbol{T} + \boldsymbol{F} \tag{120b}$$

where $\boldsymbol{F}$ is the external vector of disturbances, which is assumed constant. The quantities $\boldsymbol{p}$ and $\boldsymbol{\omega}$ are assumed with no noise. The adaptive control law is given by

$$\boldsymbol{T} = \left[\hat{L} \;\vdots\; \hat{\mathcal{J}} \;\vdots\; \hat{\boldsymbol{F}}\right]\begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{\varphi} \\ -1 \end{bmatrix} \equiv \hat{\mathcal{Q}}\boldsymbol{x} \tag{121a}$$

$$\boldsymbol{g} \equiv \begin{bmatrix} \omega_1^2 & \omega_2^2 & \omega_3^2 & \omega_1\omega_2 & \omega_2\omega_3 & \omega_1\omega_3 \end{bmatrix}^T \tag{121b}$$

$$\boldsymbol{\varphi} = -K_v\boldsymbol{\omega} - \left[\boldsymbol{\omega}\boldsymbol{\omega}^T + \left(\frac{4K_p}{1+\|\boldsymbol{p}\|^2} - \frac{\|\boldsymbol{\omega}\|^2}{2}\right)I_{3\times3}\right]\boldsymbol{p} - 4K_iB^{-1}(\boldsymbol{p})\int_0^t \boldsymbol{p}\,\mathrm{d}t \tag{121c}$$

where $\hat{\mathcal{J}}$ is the estimated inertia matrix, $\hat{\boldsymbol{F}}$ is the estimated disturbance, K_p, K_v, and K_i are scalar control gains and $\hat{L} \equiv [\hat{L}_1 \vdots \hat{L}_2]$ is a matrix of estimated inertia components, with

$$\begin{aligned} \hat{L}_1 &\equiv \begin{bmatrix} 0 & \hat{\mathcal{J}}_{23} & -\hat{\mathcal{J}}_{23} \\ -\hat{\mathcal{J}}_{13} & 0 & \hat{\mathcal{J}}_{13} \\ \hat{\mathcal{J}}_{12} & -\hat{\mathcal{J}}_{12} & 0 \end{bmatrix} \\ \hat{L}_2 &\equiv \begin{bmatrix} \hat{\mathcal{J}}_{13} & \hat{\mathcal{J}}_{33} - \hat{\mathcal{J}}_{22} & -\hat{\mathcal{J}}_{12} \\ -\hat{\mathcal{J}}_{23} & \hat{\mathcal{J}}_{12} & \hat{\mathcal{J}}_{11} - \hat{\mathcal{J}}_{33} \\ \hat{\mathcal{J}}_{22} - \hat{\mathcal{J}}_{11} & -\hat{\mathcal{J}}_{13} & \hat{\mathcal{J}}_{23} \end{bmatrix} \end{aligned} \tag{122}$$

The adaptive update law is given by

$$\dot{\hat{\mathcal{Q}}} = -\tfrac{1}{4}B^T(\boldsymbol{p})S_3\boldsymbol{e}\boldsymbol{x}^T\Gamma \tag{123}$$

where Γ is a 10×10 matrix of learning design parameters and $\boldsymbol{e}$ is given by

$$\boldsymbol{e} = \begin{bmatrix} \int_0^t(\boldsymbol{p} - \boldsymbol{p}_r)\,\mathrm{d}t \\ \boldsymbol{p} - \boldsymbol{p}_r \\ \dot{\boldsymbol{p}} - \dot{\boldsymbol{p}}_r \end{bmatrix} \tag{124}$$

where $\boldsymbol{p}_r$ and $\dot{\boldsymbol{p}}_r$ are reference trajectories. The matrix S_3 is a 3×9 submatrix of the 9×9 matrix $S = \left[S_1 \;\vdots\; S_2 \;\vdots\; S_3\right]$, which is determined by solving the following Lyapunov equation:

$$SE + E^TS = -D \tag{125}$$

where

$$E = \begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & I_{3\times3} \\ -K_iI_{3\times3} & -K_pI_{3\times3} & -K_vI_{3\times3} \end{bmatrix} \tag{126}$$

Once a matrix D is chosen, then the positive definite matrix S is determined numerically. The stability of the adaptive control law is proven using a Lyapunov analysis.

Clearly, from the definition of $\hat{L}_2$ in Eq. (122) the inertia matrix cannot be uniquely determined. Rather, the inertia matrix terms and products, using this redundant formulation to make the other parameters appear linear, are estimated such that the closed-loop dynamics assumes a prescribed linear form. The inertia matrix term adaptation is to enforce this desired linear closed-loop dynamics, not to actually identify the true inertia terms. Because of the redundancy of the inertia terms, there is an infinity of solutions of these inertia terms which will all yield the desired behavior.

XII. Conclusions

Many nonlinear filtering methods have been applied to the problem of spacecraft attitude determination in the past 25 years. This paper has provided a survey of the methods that its authors consider to be most promising. It remains the case, however, that the extended Kalman filter, especially in the form known as the multiplicative extended Kalman filter, remains the method of choice for the great majority of applications. It is a relatively simple and flexible tool with extensive heritage that can incorporate a great variety of measurements. The extended Kalman filter can fail in cases that have highly nonlinear dynamics or measurement models, or that lack a good a priori estimate of the state. Spacecraft engineers usually employ conservative designs with good initial estimates for filters, but increasingly powerful processors promote increased spacecraft autonomy, including autonomous initialization of attitude estimation filters.

Unscented filters are an attractive alternative for applications where the nonlinearities of the dynamics model or of the measurement models are severe, or when a good a priori estimate of the state is unavailable. These filters require the probability density function to be approximately Gaussian, as the central limit theorem leads one to expect in all but pathological cases, and to be at worst unimodal. Unscented filters are especially attractive when it is difficult or impossible to compute analytic partial derivatives of the dynamics or measurement models. They also have the advantage of being well suited to parallel computation. The backwards-smoothing extended Kalman filter has shown promise in situations similar to those for which unscented filters are indicated. The computational burden of the backwards-smoothing extended Kalman filter is such that it is probably not competitive with a well-designed unscented filter in most cases, though.

Particle filters are the only recourse when the density function is significantly non-Gaussian, especially if it is multimodal. Particle filters face the curse of dimensionality if more than a few parameters are to be estimated, however. Creative ways around this problem coupled with increases in computing power may make particle filters more generally useful for future spacecraft applications, but they are confined to niche applications at the present.

Some of the attitude estimation approaches presented in this paper, such as filter QUEST and recursive QUEST, are simple linear or pseudolinear filters but are suboptimal compared with the standard extended Kalman filter. These are, however, useful for spacecraft contingency designs in case of anomalies or as simple tools for analysis purposes. The orthogonal attitude filter represents the first approach to a truly nonlinear filter, but the theory is still not complete for attitude estimation. The predictive filter is useful for point-by-point estimation, but its advantages over the other filtering approaches have not been shown yet. Nonlinear observers are attractive because they usually are proven to be asymptotically stable, but due to their infancy they have not yet found widespread use on actual spacecraft. Adaptive approaches can be useful when system parameters are not known well or in the advent of spacecraft failures.

Although the new approaches surveyed here have been shown to have some advantages, it is wise to apply the old adage “if it ain’t broke don’t fix it” to the standard extended Kalman filter, which has proved its worth on a multitude of spacecraft missions. Ultimately, future mission requirements coupled with enhanced confidence in the new approaches may bring about their greater use for onboard spacecraft applications.

Acknowledgments

The authors wish to acknowledge the significant contributions and discussions on the subject matter of the following individuals: Malcolm D. Shuster from the Acme Spacecraft Company, Itzhack Bar-Itzhack and Yaakov Oshman from Technion—Israel Institute of Technology, Daniel Choukroun from Ben-Gurion University of the Negev, Hanspeter Schaub from Virginia Polytechnic Institute and State University, and Julie Thienel from the United States Naval Academy.

References

[1] Crassidis, J. L., and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, Chapman & Hall/CRC, Boca Raton, FL, 2004, Chaps. 2 and 5.

[2] Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic Press, San Diego, CA, 1970, Chaps. 4–7.

[3] Schmidt, S. F., "Kalman Filter: Its Recognition and Development for Aerospace Applications," *Journal of Guidance and Control*, Vol. 4, No. 1, 1981, pp. 4–7.

[4] Farrell, J. L., "Attitude Determination by Kalman Filter," *Automatica*, Vol. 6, No. 5, 1970, pp. 419–430.

[5] Murrell, J. W., "Precision Attitude Determination for Multimission Spacecraft," AIAA Paper 78-1248, Aug. 1978.

[6] 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, 1982, pp. 417–429.

[7] Markley, F. L., "Attitude Error Representations for Kalman Filtering," *Journal of Guidance, Control, and Dynamics*, Vol. 63, No. 2, 2003, pp. 311–317.

[8] Stuelpnagel, J., "On the Parameterization of the Three-Dimensional Rotation Group," *SIAM Review*, Vol. 6, No. 4, 1964, pp. 422–430.

[9] Shuster, M. D., "Constraint in Attitude Estimation Part I: Constrained Estimation," *Journal of the Astronautical Sciences*, Vol. 51, No. 1, 2003, pp. 51–74.

[10] Shuster, M. D., "Constraint in Attitude Estimation Part II: Unconstrained Estimation," *Journal of the Astronautical Sciences*, Vol. 51, No. 1, 2003, pp. 75–101.

[11] Markley, F. L., "Attitude Estimation or Quaternion Estimation?," *Journal of the Astronautical Sciences*, Vol. 52, Nos. 1/2, 2004, pp. 221–238.

[12] Bar-Itzhack, I. Y., and Oshman, Y., "Attitude Determination from Vector Observations: Quaternion Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 321, No. 1, 1985, pp. 128–136.

[13] Bar-Itzhack, I. Y., Deutschmann, J., and Markley, F. L., "Quaternion Normalization in Additive EKF for Spacecraft Attitude Determination," AIAA Paper 91-2706, Aug. 1991.

[14] Deutschmann, J., Markley, F. L., and Bar-Itzhack, I. Y., "Quaternion Normalization in Spacecraft Attitude Determination," *Proceedings of the Flight Mechanics/Estimation Theory Symposium, NASA-Goddard Space Flight Center, Greenbelt, MD*, CP-1992-3186, NASA Center for Aerospace Information, Hanover, MD, 1992, pp. 523–536.

[15] Psiaki, M. L., Theiler, J., Bloch, J., Ryan, S., Dill, R. W., and Warner, R. E., "ALEXIS Spacecraft Attitude Reconstruction with Thermal/ Flexible Motions Due to Launch Damage," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 5, 1997, pp. 1033–1041.

[16] Psiaki, M. L., Klatt, E. M., Kintner, P. M., and Powell, S. P., "Attitude Estimation for a Flexible Spacecraft in an Unstable Spin," *Journal of Guidance, Control, and Dynamics*, Vol. 25, No. 1, 2002, pp. 88–85.

[17] Choukroun, D., Bar-Itzhack, I. Y., and Oshman, Y., "Novel Quaternion Kalman Filter," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 42, No. 1, 2006, pp. 174–190.

[18] Psiaki, M. L., "Backward-Smoothing Extended Kalman Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 28, No. 5, 2005, pp. 885–894.

[19] Markley, F. L., Berman, N., and Shaked, U., "H_∞-Type Filter for Spacecraft Attitude Estimation," AAS Paper 93-298, May 1993.

[20] Markley, F. L., Berman, N., and Shaked, U., "Deterministic EKF-Like Estimator for Spacecraft Attitude Estimation," *Proceedings of the American Control Conference*, IEEE Publications, Piscataway, NJ, June 1994, pp. 247–251.

[21] Smith, R. H., "An H_∞-Type Filter for GPS-Based Attitude Estimation," AAS Paper 95-134, Feb. 1995.

[22] Nagpal, K. M., and Khargonekar, P. P., "Filtering and Smoothing in an H_∞ Setting," *IEEE Transactions on Automatic Control*, Vol. 36, No. 2, 1991, pp. 152–166.

[23] Haupt, G. T., Kasdin, N. J., Keiser, G. M., and Parkinson, B. W., "Optimal Recursive Iterative Algorithm for Discrete Nonlinear Least-Squares Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 19, No. 3, 1996, pp. 643–649.

[24] Shuster, M. D., and Oh, S. D., "Three-Axis Attitude Determination from Vector Observations," *Journal of Guidance and Control*, Vol. 4, No. 1, 1981, pp. 70–77.

[25] Shuster, M. D., "A Simple Kalman Filter and Smoother for Spacecraft Attitude," *Journal of the Astronautical Sciences*, Vol. 37, No. 1, 1989, pp. 89–106.

[26] Bar-Itzhack, I. Y., "REQUEST: A Recursive QUEST Algorithm for Sequential Attitude Determination," *Journal of Guidance, Control, and Dynamics*, Vol. 19, No. 5, 1996, pp. 1034–1038.

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

[28] Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. F., "A New Approach for Filtering Nonlinear Systems," *Proceedings of the American Control Conference*, IEEE Publications, Piscataway, NJ, June 1995, pp. 1628–1632.

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

[30] Nørgaard, M., Poulsen, N. K., and Ravn, O., "New Developments in State Estimation for Nonlinear Systems," *Automatica*, Vol. 36, No. 11, 2000, pp. 1627–1638.

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

[32] Doucet, A., de Freitas, N., and Gordan, N. (ed.), *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001, Chap. 1.

[33] 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, 2002, pp. 174–185.

[34] Cheng, Y., and Crassidis, J. L., "Particle Filtering for Sequential Spacecraft Attitude Estimation," AIAA Paper 04-5337, Aug. 2004.

[35] Oshman, Y., and Carmi, A., "Estimating Attitude from Vector Observations Using a Genetic Algorithm-Embedded Quaternion Particle Filter," *AIAA Guidance, Navigation, and Control Conference, Providence, RI*, AIAA, 04-5340, Aug. 2004.

[36] Crassidis, J. L., and Markley, F. L., "Predictive Filtering for Nonlinear Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 3, 1997, pp. 566–572.

[37] Daum, F. E., "New Exact Nonlinear Filters," *Bayesian Analysis of Time Series and Dynamical Models*, edited by J. C. Spall, Marcel Dekker, New York, NY, 1988, Chap. 8.

[38] Gardiner, C. W., *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Springer, Berlin, Germany, 2004, pp. 96–101.

[39] Markley, F. L., "Attitude Filtering on SO(3)," AAS, Paper 05-460, June 2005.

[40] Misawa, E. A., and Hedrick, J. K., "Nonlinear Observers: A State-of-the-Art Survey," *Journal of Dynamic Systems, Measurement and Control*, Vol. 111, No. 3, 1989, pp. 344–352.

[41] Gauthier, J. P., Hammouri, H., and Othman, S., "A Simple Observer for Nonlinear Systems Applications to Bioreactors," *IEEE Transactions on Automatic Control*, Vol. 37, No. 6, 1992, pp. 875–880.

[42] Krener, A. J., and Xiao, M., "Observers for Linearly Unobservable Nonlinear Systems," *Systems and Control Letters*, Vol. 46, No. 4, 2002, pp. 281–288.

[43] Krener, A. J., and Kang, W., "Locally Convergent Nonlinear Observers," *Systems and Control Letters*, Vol. 42, No. 1, 2003, pp. 155–177.

[44] Michalska, H., and Mayne, D. Q., "Moving Horizon Observers and Observer-Based Control," *IEEE Transactions on Automatic Control*, Vol. 40, No. 6, 1995, pp. 995–1006.

[45] Rao, C. V., Rawlings, J. B., and Mayne, D. Q., "Constrained State Estimation for Nonlinear Discrete-Time Systems: Stability and Moving Horizon Approximations," *IEEE Transactions on Automatic Control*, Vol. 48, No. 2, 2003, pp. 246–258.

[46] Gong, Q., Ross, I. M., and Kang, W., "A Pseudospectral Observer for Nonlinear Systems," AIAA Paper 05-5845, Aug. 2005.

[47] Luk'yanov, A. G., Dodds, S. D., and Vittek, J., "Observer-Based Attitude Control in the Sliding Mode," *Proceedings of the Third International Conference on Dynamics and Control of Structures in Space*, Computational Mechanics, Inc., Billeria, MA, June 1996, pp. 639–671.

[48] Nicosia, S., and Tomei, P., "Nonlinear Observer and Output Feedback Attitude Control of Spacecraft," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 4, 1996, pp. 970–977.

[49] McDuffie, J. H., and Shtessel, Y. B., "A Sliding Mode Controller and Observer for Satellite Attitude Control," AIAA Paper 97-3755, Aug. 1997.

[50] Algrain, M. C., and Lee, M., "Discrete-Time Nonlinear Observer Design for Reproducing Angular Rates Along the Third-Axis of a Spinning Spacecraft Using Only Two-Axis Measurements," *Proceedings of the Aerospace and Electronics Conference*, Vol. 2, IEEE Publications, Piscataway, NJ, July 1997, pp. 638–645.

[51] Bošković, J. D., Li, S.-M., and Mehra, R. K., "A Globally Stable Scheme for Spacecraft Control in the Presence of Sensor Bias," *Proceedings of the IEEE Aerospace Conference*, Vol. 3, IEEE Publications, Piscataway, NJ, Mar. 2000, pp. 505–511.

[52] Thienel, J. K., and Sanner, R. M., "A Coupled Nonlinear Spacecraft Attitude Controller and Observer with an Unknown Constant Gyro Bias and Gyro Noise," *IEEE Transactions on Automatic Control*, Vol. 48, No. 11, 2003, pp. 2011–2015.

[53] Mehra, R. K., "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, Vol. 15, No. 2, 1970, pp. 175–184.

[54] Ma, Z., and Ng, A., "Spacecraft Attitude Determination by Adaptive Kalman Filtering," AIAA Paper 02-4460, Aug. 2002.

[55] Lam, Q. M., and Wu, A., "Enhanced Precision Attitude Determination Algorithms," *Proceedings of the IEEE Aerospace Conference*, Vol. 1, IEEE Publications, Piscataway, NJ, Mar. 1998, pp. 61–68.

[56] Mehra, R., Seereeram, S., Bayard, D., and Hadaegh, F., "Adaptive Kalman Filtering, Failure Detection and Identification for Spacecraft Attitude Estimation," *Proceedings of the 4th IEEE Conference on Control Applications*, IEEE Publications, Piscataway, NJ, Sept. 1995, pp. 176–181.

[57] Rapoport, I., and Oshman, Y., "Optimal Filtering in the Presence of Faulty Measurement Biases," *Proceedings of 41st IEEE Conference on Decision and Control*, IEEE Publications, Piscataway, NJ, Dec. 2002, pp. 2236–2241.

[58] Carter, M. T., Vadali, S. R., and Chamitoff, G. E., "Parameter Identification for the International Space Station Using Nonlinear Momentum Management Control," AIAA Paper 97-3524, Aug. 1997.

[59] Kim, J.-W., Crassidis, J. L., Vadali, S. R., and Dershowitz, A. L., "International Space Station Leak Localization Using Vent Torque Estimation," IAC Paper 04-A.4.10, Oct. 2004.

[60] Psiaki, M. L., "Estimation of the Parameters of Spacecraft's Attitude Dynamics Model Using Flight Data," *Proceedings of the Flight Mechanics Symposium*, CP-2003-212246, NASA Goddard Space Flight Center, Greenbelt, MD, 2003, Session 5, Paper 3.

[61] Kim, I., Kim, J., and Kim, Y., "Angular Rate Estimator Using Disturbance Accommodation Technique," AIAA Paper 02-4829, Aug. 2002.

[62] Schaub, H., Akella, M. R., and Junkins, J. L., "Adaptive Control of Nonlinear Attitude Motions Realizing Linear Closed Loop Dynamics," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 1, 2001, pp. 95–100.

[63] Costic, B. T., Dawson, D. M., de Queiroz, M. S., and Kapila, V., "Quaternion-Based Adaptive Attitude Tracking Controller Without Velocity Measurements," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 6, 2001, pp. 1214–1222.

[64] Tanygin, S., "Generalization of Adaptive Attitude Tracking," AIAA Paper 02-4833, Aug. 2002.

[65] Goldstein, H., *Classical Mechanics*, 2nd ed., Addison-Wesley Publishing Company, Reading, MA, 1980, Chap. 4.

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

[67] Hamilton, W. R., *Elements of Quaternions*, Longmans, Green and Co., London, England, 1866.

[68] Shepperd, S. W., "Quaternion from Rotation Matrix," *Journal of Guidance and Control*, Vol. 1, No. 3, 1978, pp. 223–224.

[69] Kuipers, J. B., *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace, and Virtual Reality*, Princeton University Press, Princeton, NJ, 1999.

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

[71] Pittelkau, M. E., "Kalman Filtering for Spacecraft System Alignment Calibration," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 6, 2001, pp. 1187–1195.

[72] Pandiyan, R., Solaiappan, A., and Malik, N., "A One Step Batch Filter for Estimating Gyroscope Calibration Parameters Using Star Vectors," AIAA Paper 04-4858, Aug. 2004.

[73] Crassidis, J. L., "Sigma-Point Kalman Filtering for Integrated GPS and Inertial Navigation," AIAA Paper 05-6052, Aug. 2005.

[74] Shuster, M. D., "Kalman Filtering of Spacecraft Attitude and the QUEST Model," *Journal of the Astronautical Sciences*, Vol. 38, No. 3, 1990, pp. 377–393.

[75] Sedlak, J., and Chu, D., "Kalman Filter Estimation of Attitude and Gyro Bias with the QUEST Observation Model," AAS Paper 93-297, May 1993.

[76] Cheng, Y., Crassidis, J. L., and Markley, F. L., "Attitude Estimation for Large Field-of-View Sensors," AAS Paper 05-462, June 2005.

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

[78] Varotto, S. E. C., Orlando, V., and Lopes, R. V. F., "Um Procedimento para Determinação da Atitude de Satélites Artificiais Utilizando Técnicas de Estimação Ótima Estática e Dinâmica (A Procedure for Attitude Determination of Artificial Satellites Using Static and Dynamic Optimal Estimation Techniques)," *Proceedings of the 6th Brazilian Automatic Control Conference*, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, 1986, pp. 946–951.

[79] Idan, M., "Estimation of Rodrigues Parameters from Vector Observations," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 32, No. 2, 1996, pp. 578–586.

[80] Crassidis, J. L., and Markley, F. L., "Attitude Estimation Using Modified Rodrigues Parameters," *Proceedings of the Flight Mechanics/Estimation Theory Symposium*, CP-1996-3333, NASA Goddard Space Flight Center, Greenbelt, MD, 1996, pp. 71–83.

[81] 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.

[82] Toda, N. F., Heiss, J. L., and Schlee, F. H., "SPARS: The System, Algorithm, and Test Results," Aerospace Corp. Rept. TR-0066 (5306)-12, 1969, pp. 361–370.

[83] Yong, K., and Headley, R. P., "Real Time Precision Attitude Determination System (RETPAD) for Highly Maneuverable Spacecrafts," AIAA Paper 78-1246, Aug. 1978.

[84] Pittelkau, M. E., "Rotation Vector Attitude Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 26, No. 6, 2003, pp. 855–860.

[85] Gray, C. W., "Star Tracker/IRU Attitude Determination Filters," AAS Paper 01-039, Feb. 2001.

[86] Thompson, I. C., and Quasius, G. R., "Attitude Determination for the P80-1 Satellite," AIAA Paper 80-001, Feb. 1980.

[87] Gai, E., Daly, K., Harrison, J., and Lemos, L., "Star-Sensor-Based Satellite Attitude/Attitude Rate Estimator," *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 5, 1985, pp. 560–565.

[88] Paulson, D. C., Jackson, D. B., and Brown, C. D., "SPARS Algorithms and Simulation Results," Aerospace Corp. Rept. TR-0066 (5306)-12, 1969, pp. 293–317.

[89] Oshman, Y., and Markley, F. L., "Sequential Attitude and Attitude-Rate Estimation Using Integrated-Rate Parameters," *Journal of Guidance, Control, and Dynamics*, Vol. 22, No. 3, 1999, pp. 385–394.

[90] Vathsal, S., "Spacecraft Attitude Determination Using a Second-Order Nonlinear Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 10, No. 6, 1987, pp. 559–566.

[91] Bar-Itzhack, I. Y., and Reiner, J., "Recursive Attitude Determination from Vector Observations: DCM Identification," *Journal of Guidance, Control, and Dynamics*, Vol. 7, No. 1, 1984, pp. 51–65.

[92] Pittelkau, M. E., "An Analysis of the Quaternion Attitude Determination Filter," *Journal of the Astronautical Sciences*, Vol. 51, No. 1, 2003, pp. 103–120.

[93] Gelb, A. (ed.), *Applied Optimal Estimation*, The MIT Press, Cambridge, MA, 1974, pp. 190–191.

[94] Wahba, G., "A Least-Squares Estimate of Satellite Attitude," *SIAM Review*, Vol. 7, No. 3, 1965, pp. 409.

[95] Golub, G. H., and Van Loan, C. F., *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996, pp. 140–149, 601.

[96] Markley, F. L., and Mortari, D., "Quaternion Attitude Estimation Using Vector Observations," *Journal of the Astronautical Sciences*, Vol. 48, No. 2/3, 2000, pp. 359–380.

[97] Lerner, G. M., "q Method," *Spacecraft Attitude Determination and Control*, edited by J. R. Wertz, Kluwer Academic Publishers, The Netherlands, 1978, pp. 426–428.

[98] Kim, S.-G., Crassidis, J. L., Cheng, Y., Fosbury, A. M., and Junkins, J. L., "Kalman Filtering for Relative Spacecraft Attitude and Position Estimation," *Journal of Guidance, Control, and Dynamics* (to be published).

[99] Choukroun, D., Bar-Itzhack, I. Y., and Oshman, Y., "Optimal-REQUEST Algorithm for Attitude Determination," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 3, 2000, pp. 418–425.

[100] Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation*, Academic International Press, Orlando, FL, 1977, pp. 57–67, 69–76, 115–122.

[101] Kasdin, N. J., "Satellite Quaternion Estimation from Vector Measurements with the Two-Step Optimal Estimator," AAS Paper 02-002, Feb. 2002.

[102] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, New York, NY, 2001, Chap. 5.

[103] Daum, F., and Huang, J., "Curse of Dimensionality and Particle Filters," *Proceedings of the IEEE Aerospace Conference*, Vol. 4, IEEE Publications, Piscataway, NJ, Mar. 2003, pp. 1979–1993.

[104] Ristic, B., Arulampalam, S., and Gordon, N., *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, Boston, MA, 2004, Chap. 3.

[105] Crisan, D., and Doucet, A., "A Survey on Convergence Results on Particle Filtering Methods for Practitioners," *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, 2002, pp. 736–746.

[106] Oshman, Y., and Carmi, A., "Spacecraft Attitude Estimation from Vector Observations Using a Fast Particle Filter," AAS Paper 04-141, Feb. 2004.

[107] Oshman, Y., and Carmi, A., "Adaptive Estimation of Spacecraft Attitude from Vector Observations Using a Quaternion Particle Filter," AAS Paper 05-464, June 2005.

[108] Moakher, M., "Means and Averaging in the Group of Rotations," *SIAM Journal on Matrix Analysis and Applications*, Vol. 24, No. 1, 2002, pp. 1–16.

[109] Shuster, M. D., "Uniform Attitude Probability Distributions," *Journal of the Astronautical Sciences*, Vol. 51, No. 4, 2003, pp. 451–475.

[110] Lo, J. T.-H., "Optimal Estimation and Detection for Rotational Processes," *Proceedings of the Flight Mechanics/Estimation Theory Symposium, NASA-Goddard Space Flight Center, Greenbelt, MD*, CP-1976-2023, NASA Center for Aerospace Information, Hanover, MD, 1976, pp. 83–87.

[111] Lo, J. T.-H., and Eshleman, L. R., "Exponential Fourier Densities on SO(3) and Optimal Estimation and Detection for Rotational Processes," *SIAM Journal on Applied Mathematics*, Vol. 36, No. 1, 1979, pp. 73–82.

[112] Lo, J. T.-H., "Optimal Estimation for the Satellite Attitude Using Star Tracker Measurements," *Automatica*, Vol. 22, No. 4, 1986, pp. 477–482.

[113] Biedenharn, L. C., and Louck, J. D., *Angular Momentum in Quantum Physics*, Vol. 8, Encyclopedia of Mathematics and its Applications, Addison–Wesley, Reading, MA, 1981, Chap. 3.

[114] Hunt, L. R., Luksic, M., and Su, R., "Exact Linearizations of Input-Output Systems," *International Journal of Control*, Vol. 43, No. 1, 1986, pp. 247–255.

[115] Lu, P., "Nonlinear Predictive Controllers for Continuous Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 17, No. 3, 1994, pp. 553–560.

[116] Crassidis, J. L., Lightsey, E. G., and Markley, F. L., "Efficient and Optimal Attitude Determination Using Recursive Global Positioning System Signal Operations," *Journal of Guidance, Control, and Dynamics*, Vol. 22, No. 2, 1999, pp. 193–201.

[117] Crassidis, J. L., and Markley, F. L., "Predictive Filtering for Attitude Estimation Without Rate Sensors," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 3, 1997, pp. 522–527.

[118] Crassidis, J. L., Alonso, R., and Junkins, J. L., "Optimal Attitude and Position Determination from Line-of-Sight Measurements," *Journal of the Astronautical Sciences*, Vol. 48, No. 2/3, 2000, pp. 391–408.

[119] Thienel, J. K., and Sanner, R. M., "Hubble Space Telescope Angular Velocity Estimation During the Robotic Servicing Mission," *Journal of Guidance, Control, and Dynamics* (to be published).

[120] Khalil, H., *Nonlinear Systems*, Prentice Hall, Upper Saddle River, NJ, 1996, pp. 115–116.

[121] Schaub, H., and Junkins, J. L., *Analytical Mechanics of Aerospace Systems*, American Institute of Aeronautics and Astronautics, Inc., New York, NY, 2003, pp. 136–138.