Nonlinear Attitude Filtering Methods

F. Landis Markley*

NASA Goddard Space Flight Center, Greenbelt, MD 20771

John L. Crassidis,† Yang Cheng‡

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

This paper provides a survey of modern nonlinear filtering methods for attitude estimation. Early applications relied mostly on the extended Kalman filter for attitude estimation. Since these applications, several new approaches have been developed that have proven to be superior to the extended Kalman filter. Several of these approaches maintain the basic structure of the extended Kalman filter, but employ various modifications in order to provide better convergence or improve other performance characteristics. Examples of such approaches include: filter QUEST, extended QUEST, the super-iterated extended Kalman filter, the interlaced extended Kalman filter, and the second-order Kalman filter. Filters that propagate and update a discrete set of sigma points rather than using linearized equations for the mean and covariance are also reviewed. A two-step approach is discussed with a first-step state that linearizes the measurement model and an iterative second step to recover the desired attitude states. These approaches are all based on the Gaussian assumption that the probability density function is adequately specified by its mean and covariance. Other approaches that do not require this assumption are reviewed, including particle filters and a Bayesian filter based on a non-Gaussian, finite-parameter probability density function on SO(3). Finally, the predictive filter, nonlinear observers and adaptive approaches are shown. The strengths and weaknesses of the various approaches are discussed.

I. Introduction

The extended Kalman filter[1–7] (EKF) is the workhorse of real-time spacecraft attitude estimation. Since the group SO(3) of rotation matrices 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] In particular, the backwards-smoothing EKF solves a nonlinear smoothing problem for the current and past sample intervals using iterative numerical techniques. The numerical iteration retains all of the nonlinearities of a fixed number of stages that precede the terminal stage of interest, and it processes information from earlier stages in an approximate manner.[18] Deterministic EKF-like estimators are closely related to H_∞ control design.[22] Usually an upper bound is derived first, and then the bound is minimized based on

---

*Aerospace Engineer, Guidance, Navigation and Control Systems Engineering Branch. Email: Landis.Markley@nasa.gov. Fellow AIAA.

†Associate Professor, Department of Mechanical & Aerospace Engineering. Email: johnc@eng.buffalo.edu. Associate Fellow AIAA.

‡Postdoctoral Research Fellow, Department of Mechanical & Aerospace Engineering. Email: cheng3@eng.buffalo.edu. Member AIAA.

approaches such as Riccati equations and linear matrix inequalities. The nonlinear problem is much more difficult, which requires the solution of the Hamilton-Jacobi-Isaacs partial differential inequality. For attitude estimation, small-error approximations are used to developed a filter.[19,20]

Other designs use various assumptions to derive simplified filters. These generally provide suboptimal performance characteristics in relation to the EKF, but involve linear or pseudo-linear 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 deterministic solution of the attitude, e.g. methods that are based on the QUEST attitude determination solution.[23] Simple filter designs based on QUEST include filter QUEST[24] and recursive QUEST.[25] A more complicated but far more robust approach, called extended QUEST,[26] uses a full nonlinear propagation along with a novel measurement update. This approach can be used to estimate attitude and additional parameters as well. Interlaced filters that replace a nonlinear filter with two or more linear filters have been used for rate estimation, but not for attitude estimation.[27–29]

Several new alternatives to the standard EKF have been recently introduced, such as sigma point or unscented filters[30–33] and particle filters.[34–37] Unscented filters (UFs) have been shown to exhibit several advantages over the EKF, including: 1) the expected error is lower than the EKF, 2) they can be applied to non-differentiable functions, 3) no Jacobian matrix calculations are required, and 4) they provide higher-order expansions than the standard EKF. Unscented filters work 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. They typically use the standard Kalman form in the post-update, but use a different propagation of the covariance and pre-measurement update with no local iterations. The attitude estimation UF derived in Ref. 33 is based on a quaternion representation of the attitude kinematics. However, straightforward implementation of the standard UF equations derives a predicted quaternion mean from an averaged sum of quaternions. Therefore, no guarantees can be made that the resulting quaternion will have unit norm. This was overcome by using a generalized unconstrained three-component vector to represent the attitude error-quaternion, leading to an unconstrained formulation in the UF design.

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 the sufficient statistics of a Gaussian distribution, they are not sufficient to represent a general probability distribution. When UF methods are applied to strongly nonlinear and non-Gaussian estimation problems, where the *a posteriori* distribution of the state vector may be multi-peaked, heavily-tailed, or skewed, desired performance characteristics may not be obtained. This may be overcome by using particle filters[34–37] (PFs). Like other approximate approaches to optimal filtering, the ultimate objective of PFs is to construct the *a posteriori* probability density function (PDF) of the state vector, or the PDF of the state vector conditional on all the available measurements. However, the approximation of PFs is vastly different from that of conventional nonlinear filters. The central idea of a PF approximation is to represent a continuous distribution of interest by a finite (but large) number of weighted random samples of the state vector, or particles. A PF does not assume the *a posteriori* distribution of the state vector to be a Gaussian distribution or any other distribution of known form. In principle, it can estimate probability distributions of arbitrary form and solve any nonlinear and/or non-Gaussian system. Reference 36 presents a PF for attitude estimation based on the bootstrap filter.[35]

The optimal solution of the nonlinear estimation problem requires the propagation of the conditional PDF of the state given the observation history.[37] All practical nonlinear filters are approximations to this ideal. Exact finite dimensional filters[38] can be found that solve some nonlinear problems by using the Fokker-Planck equation[2,39] to propagate a non-Gaussian PDF between measurements and Bayes' formula[1,2] to incorporate measurement information. A recently proposed filter[40] follows this pattern, but does not solve the nonlinear attitude filtering problem exactly. We refer to it as an orthogonal filter, because it represents the attitude by an orthogonal rotation matrix, rather than by some parameterization of the rotation matrix. The PDF is a non-Gaussian function defined on the Cartesian product of SO(3), the group of rotation matrices, and the Euclidean space $\mathbb{R}^N$ of bias parameters. This filter entirely avoids questions about singularities of representations or covariance matrices arising in EKFs[6–10] and UFs,[33] and has the additional advantage of providing a consistent initialization for a completely unknown initial attitude, owing to the fact that SO(3) is a compact space.

Nonlinear observers often exhibit global convergence, which is to say that they can converge from any initial guess.[41] Several applications of observers for attitude control have been proposed Refs. 42–48. A

American Institute of Aeronautics and Astronautics

nonlinear observer and controller using only measurements of roll, pitch and yaw has been developed in Ref. 43. Local asymptotic stability is ensured under mild hypotheses. A globally convergent, nonlinear full-order observer using quaternions and Euler's equations for the dynamics has been derived in Ref. 44. The observer structure contains a discontinuous term, which is often associated with "sliding mode" observers. The error-quaternion is defined using a multiplicative approach and stability is proven using a Lyapunov function. A simpler, robust smoothed sliding mode observer that avoids quaternion error differentiation noise and eliminates the necessity of measuring angular rate is derived in Ref. 45. Although an additive approach is used to define the quaternion error, global stability is still provided through a Lyapunov function. Algrain and Lee develop a nonlinear observer to estimate angular rates along the third axis of a spinning spacecraft using only two-axis measurements.[46] A pseudo-linear model is developed by decomposing the nonlinear system into linear and nonlinear parts. Bošković, Li and Mehra use angular rate measurements with quaternion kinematics to derive a nonlinear bias observer, which is coupled with an adaptive sliding mode controller.[47] Stability is proven as long as the attitude never passes through ±180 rotations. Thienel and Sanner develop an exponentially convergent nonlinear observer given a constant gyro bias with identification of the bias proven through a persistency of excitation argument.[48] An analysis is also shown that includes gyro noise.

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. In practice "tuning" a Kalman filter can be arduous and very time-consuming. Usually, the measurement-error covariance is fairly well known, derived from statistical inferences of the hardware sensing device. However, the process noise covariance is usually not well known and is often derived from experiences gained by the design engineer based on intimate knowledge of the particular system. The approach is based on "residual whitening."[49] Unfortunately, most noise adaptive techniques are applicable only for linear systems,[1] which creates problems for attitude estimation due to the nonlinear equations involved. Still, it is possible to use these techniques with linearized equations, as demonstrated in Ref. 50. Lam and Wu further develop adaptive filters that address both colored and white noise statistics.[51] The former noise is identified using a non-parametric neural network approach, while the latter noise is identified using an α-β filter. An adaptive filter is also proposed in Ref. 17 to account for inaccuracy in the knowledge of the process noise statistical model, which uses a linear pseudo-measurement model. Other adaptive approaches use adaptive methods for fault tolerant estimation purposes.[52,53] 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 Refs. 54–56. A disturbance accommodation technique that models the unknown disturbance angular rate using a power set of time as basis functions is shown in Ref. 57. Nonlinear adaptive techniques are similar to nonlinear observers in that they usually provide global stability proofs that guarantee convergence of the estimated parameters.[58–60]

This paper will review the basic assumptions of these filters, 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. The 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 non-moving, 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, i.e. its inverse is given by its transpose, and proper, i.e. its determinant is +1.[61] 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

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

American Institute of Aeronautics and Astronautics

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

Several parameterizations of the attitude are possible.[62] 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, as is discussed later. For modern-day applications, i.e. since the early 1980s, the quaternion[63] has been the most widely used attitude parameterization. The quaternion is a four-dimensional vector, defined as

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

with

$$\varrho \equiv [q_1 \ q_2 \ q_3]^T = \mathbf{e}\sin(\vartheta/2) \tag{3a}$$

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

where $\mathbf{e}$ is the unit Euler axis and ϑ is the rotation angle. Since 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 $\mathbf{q}^T\mathbf{q} = 1$. The attitude matrix is related to the quaternion by

$$A(\mathbf{q}) = \left(q_4^2 - ||\varrho||^2\right) I_{3\times3} + 2\varrho\,\varrho^T - 2q_4[\varrho\times] = \Xi^T(\mathbf{q})\Psi(\mathbf{q}) \tag{4}$$

where $I_{3\times3}$ is a 3×3 identity matrix and

$$\Xi(\mathbf{q}) \equiv \begin{bmatrix} q_4 I_{3\times3} + [\varrho\times] \\ -\varrho^T \end{bmatrix} \tag{5a}$$

$$\Psi(\mathbf{q}) \equiv \begin{bmatrix} q_4 I_{3\times3} - [\varrho\times] \\ -\varrho^T \end{bmatrix} \tag{5b}$$

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

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

For small angles the vector part of the quaternion is approximately equal to half angles,[62] which will be used later.

The quaternion kinematics equation is given by

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

where ω is the three-component angular rate vector and

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

A useful identity is given by

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

where

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

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 we adopt the convention of Ref. 6, who multiply the quaternions in the same order as the attitude matrix multiplication, in contrast to the usual convention established by Hamiliton.[63] Suppose we wish to perform a successive rotation. This can be written using

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

The composition of the quaternions is bilinear, with

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

Also, the inverse quaternion is defined by

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

Note that $\mathbf{q} \otimes \mathbf{q}^{-1} = [0 \ 0 \ 0 \ 1]^T$, which is the identity quaternion. A computationally efficient algorithm to extract the quaternion from the attitude matrix is given in Ref. 64. A more thorough review of the quaternion parameterization, as well as other parameterizations, can be found in the survey paper by Shuster[62] and in the book by Kuipers.[65]

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[66]

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

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

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

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

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

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.[67,68] For simulation purposes, discrete-time gyro measurements can be generated using the following equations:[69]

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

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

where the subscript k denotes the kth time-step, Δt is the gyro sampling interval, and $\mathbf{N}_v$ and $\mathbf{N}_u$ 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[62] 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] so we are faced with the alternatives of using an attitude representation that is either singular or redundant. The various strategies to face or evade this dilemma can be divided into three general classes, which we will refer to as the minimal representation EKF, the multiplicative EKF (MEKF) and the additive EKF (AEKF). Reference 6 presents an overview of Kalman filtering for spacecraft attitude estimation, emphasizing the quaternion representation, with a complete list of references through 1981. 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).[70, 71] The UVF employs also a very useful approximation to the errors in the unit vector measurements. In particular, Shuster and Oh[23] have shown that nearly all the probability of the errors is concentrated on a very small area about the direction of $A\mathbf{r}$, so the sphere containing that point can be approximated by a tangent plane, characterized by

$$\tilde{\mathbf{b}} = A\mathbf{r} + \boldsymbol{v}, \quad \boldsymbol{v}^T A\mathbf{r} = 0 \tag{17}$$

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

$$E\{\boldsymbol{v}\} = \mathbf{0} \tag{18a}$$

$$R \equiv E\{\boldsymbol{v}\boldsymbol{v}^T\} = \sigma^2 \left[I_{3\times3} - (A\mathbf{r})(A\mathbf{r})^T \right] \tag{18b}$$

Equation (18b) is known as the QUEST measurement model.[23, 70, 71] This model is quite accurate for small field-of-view sensors. The approximations in this error model are discussed in Ref. 72 and 73. 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

$$R = \sigma^2 I_{3\times3} \tag{19}$$

gives equivalent results.[70] The QUEST measurement model has been expanded for large field-of-views in Ref. 73.

Using QUEST[23] or an equivalent quaternion estimator as a data compressor simplifies the interface even further. This is especially useful since 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, 70, 74]

B. Minimal Representation EKF

The rotation group has three dimensions, so the most straightforward implementation of an EKF employs a three-dimensional parameterization of the attitude. The earliest 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 degrees, 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[75] and the MRPs have been developed for this application.[76] The MRPs are nonsingular for rotations less than 360 degrees, and the singularity can be avoided by changing to a "shadow set" of parameters.[77] 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, 78-80] In this case the product is

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

where $\hat{\mathbf{q}}$ is the unit estimated quaternion and $\delta\mathbf{q}(\boldsymbol{\phi})$ is a unit quaternion representing the rotation from $\hat{\mathbf{q}}$ to the true attitude $\mathbf{q}$, 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.[81-83]

It is also possible to represent the reference attitude by an estimated attitude matrix $\hat{A}$ rather than by an estimated quaternion.[7,81,84,85] 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, while restoring the orthogonality of $\hat{A}$ is nontrivial. Gray has argued that this argument is not compelling if reasonable computational care is taken.[81]

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

Several choices for ϕ have been used,[7] including the vector of infinitesimal rotation angles,[80] two times the vector part of the quaternion,[6] two times the vector of Rodrigues parameters,[7] four times the vector of MRPs, or the integrated rate parameters (IRPs).[85] All these choices have the small-angle approximation

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

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 Precise Attitude Reference System (SPARS) in 1969,[78,84] 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 Refs. 6 and 7. The latter reference discusses the extension of the MEKF to a second-order filter, following earlier work by Vathsal.[86]

D. Additive EKF

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.[87] We will only discuss the quaternion EKF.

The quaternion AEKF relaxes the quaternion normalization condition and treats the four components of the quaternion as independent parameters. It defines the estimate $\hat{q}$ and error Δq by

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

This means that

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

where $\bar{y}$ denotes the measurement vector. The equality in Eq. (23) is valid only if Δq is identically zero. Equation (23) shows that if the random variable 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 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 $q/||q||$ in the homogenous quadratic form, giving

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

The subscript R identifies this as the ray representation model, since 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 (LOM) model.[13,14] The ray representation form of the AEKF, which is effectively equivalent to the MEKF in the limit of continuous measurements,[72] has been applied to attitude estimation of the ALEXIS and CAPER spacecraft.[15,16]

Choukroun, Bar-Itzhack, and Oshman 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,

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

where $\mathbf{v}$ 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,[72] but has subtle differences for discrete measurements.

The relative merits of the AEKF and the MEKF have been discussed at length.[9,10,72,88] One disadvantage of the 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 EKF

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

$$J_k = \frac{1}{2} \sum_{i=0}^{k-1} \left\{ [\tilde{\mathbf{y}}_{i+1} - \mathbf{h}_{i+1}(\mathbf{x}_{i+1})]^T R_{i+1}^{-1} [\tilde{\mathbf{y}}_{i+1} - \mathbf{h}_{i+1}(\mathbf{x}_{i+1})] + \mathbf{w}_i^T Q_i^{-1} \mathbf{w}_i \right\}$$
$$+ \frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T P_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) \tag{26}$$

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

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

The process noise covariance is Q_i, the measurement noise covariance is R_i, $\tilde{\mathbf{y}}_{i+1}$ is the measurement at time t_{i+1}, $\mathbf{h}_{i+1}(\mathbf{x}_{i+1})$ is the nonlinear measurement model, and $\hat{\mathbf{x}}_0$ 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 $\hat{\mathbf{x}}_i$ for $i < k$ when $\mathbf{x}_k$ is optimized in the kth step. The iterated EKF improves upon the EKF by iterating the nonlinear measurement update equation for $\hat{\mathbf{x}}_k$, re-linearizing about the updated state estimate at each iteration,[1,89] but it does not explicitly recompute the values of $\hat{\mathbf{x}}_i$ 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 EKF (BSEKF), or super-iterated 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 $\mathbf{x}_k$ along with $\mathbf{x}_i$ and $\mathbf{w}_i$ for $i = k - m(k), \ldots, k-1$ to minimize the loss function

$$J_k = \frac{1}{2} \sum_{i=k-m(k)}^{k-1} \left\{ [\tilde{\mathbf{y}}_{i+1} - \mathbf{h}_{i+1}(\mathbf{x}_{i+1})]^T R_{i+1}^{-1} [\tilde{\mathbf{y}}_{i+1} - \mathbf{h}_{i+1}(\mathbf{x}_{i+1})] + \mathbf{w}_i^T Q_i^{-1} \mathbf{w}_i \right\}$$
$$+ \frac{1}{2}[\mathbf{x}_{k-m(k)} - \hat{\mathbf{x}}_{k-m(k)}^*]^T [P_{k-m(k)}^*]^{-1} [\mathbf{x}_{k-m(k)} - \hat{\mathbf{x}}_{k-m(k)}^*] \tag{28}$$

subject to the dynamics equation

$$\mathbf{x}_{i+1} = \mathbf{f}_i(\mathbf{x}_i, \mathbf{w}_i) \quad \text{for} \quad i = k - m(k), \ldots, k-1 \tag{29}$$

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

American Institute of Aeronautics and Astronautics

to the loss function $J_{k-m(k)}$ for fixed $\mathbf{x}_{k-m(k)}$ optimized over all the $\mathbf{x}_i$ and $\mathbf{w}_i$ 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 BSEKF solves nonlinear least-squares problem of Eqs. (28) and (29) using a guarded Gauss-Newton method,[90] which works with guesses and increments of $\mathbf{x}_{k-m(k)}$, $\mathbf{w}_{k-m(k)}$, $\mathbf{w}_{k-m(k)+1}$, $\mathbf{w}_{k-m(k)+2}$, $\cdots$, and $\mathbf{w}_{k-1}$. Basing the initial guesses $\mathbf{x}^0_{k-m(k)}$, $\mathbf{w}^0_{k-m(k)}$, $\mathbf{w}^0_{k-m(k)+1}$, $\mathbf{w}^0_{k-m(k)+2}$, $\cdots$, and $\mathbf{w}^0_{k-1}$ for the kth measurement update on the converged values of the previous update minimizes the number of Gauss-Newton iterations required. The state estimates for the jth Gauss-Newton iteration are propagated by

$$\mathbf{x}^j_{i+1} = \mathbf{f}_i(\mathbf{x}^j_i, \mathbf{w}^j_i) \quad \text{for} \quad i = k - m(k), \ldots, k - 1 \tag{30}$$

The loss function J^j_k for this iteration and the measurement residuals and partial derivative matrices

$$\Delta \mathbf{y}_{i+1} = \tilde{\mathbf{y}}_{i+1} - \mathbf{h}^j_{i+1}(\mathbf{x}_{i+1}), \quad \Phi_i = \partial \mathbf{f}_i / \partial \mathbf{x}_i, \quad \Gamma_i = \partial \mathbf{f}_i / \partial \mathbf{w}_i, \quad \text{and} \quad H_i = \partial \mathbf{h}_i / \partial \mathbf{x}_i \tag{31}$$

are computed , and $\Delta \mathbf{x}_{k-m(k)}$, $\Delta \mathbf{w}_{k-m(k)}$, $\Delta \mathbf{w}_{k-m(k)+1}$, $\Delta \mathbf{w}_{k-m(k)+2}$, $\ldots$, and $\Delta \mathbf{w}_{k-1}$ are found to minimize the loss function

$$J_k = \frac{1}{2} \sum_{i=k-m(k)}^{k-1} \left[(\Delta \mathbf{y}_{i+1} - H_{i+1} \Delta \mathbf{x}_{i+1})^T R^{-1}_{i+1} (\Delta \mathbf{y}_{i+1} - H_{i+1} \Delta \mathbf{x}_{i+1}) + (\mathbf{w}^j_i + \Delta \mathbf{w}_i)^T Q^{-1}_i (\mathbf{w}^j_i + \Delta \mathbf{w}_i) \right]$$
$$+ \frac{1}{2} [\mathbf{x}^j_{k-m(k)} + \Delta \mathbf{x}_{k-m(k)} - \hat{\mathbf{x}}^*_{k-m(k)}]^T [P^*_{k-m(k)}]^{-1} [\mathbf{x}^j_{k-m(k)} + \Delta \mathbf{x}_{k-m(k)} - \hat{\mathbf{x}}^*_{k-m(k)}] \tag{32}$$

subject to the dynamics equation

$$\Delta \mathbf{x}_{i+1} = \Phi_i \Delta \mathbf{x}_i + \Gamma_i \Delta \mathbf{w}_i \tag{33}$$

This linear fixed-interval smoothing problem is solved using a square-root information matrix method that has a forward filtering pass followed by a backwards smoothing pass.[91] Now the guesses are updated by

$$\mathbf{x}^{j+1}_{k-m(k)} = \mathbf{x}^j_{k-m(k)} + \alpha \, \Delta \mathbf{x}_{k-m(k)} \tag{34a}$$

$$\mathbf{w}^{j+1}_i = \mathbf{w}^j_i + \alpha \, \Delta \mathbf{w}_i \quad \text{for} \quad i = k - m(k), \ldots, k - 1 \tag{34b}$$

$$\mathbf{x}^{j+1}_{i+1} = \mathbf{f}_i(\mathbf{x}^{j+1}_i, \mathbf{w}^{j+1}_i) \quad \text{for} \quad i = k - m(k), \ldots, k - 1 \tag{34c}$$

for $\alpha = 1$, and the updated loss function J^{j+1}_k is calculated. If the loss function has decreased, the next Gauss-Newton iteration is computed. If the loss function has increased, the value of α is halved and the updates of Eq. (34) are recomputed. The Gauss-Newton iterations are repeated until either termination criteria are satisfied or a limiting number of iterations is exceeded.

The quantities $\hat{\mathbf{x}}^*_{k-m(k)}$ and $P^*_{k-m(k)}$ must be computed to insert into Eq. (28) to execute the next time step. Contrary to appearances, these are not the *a posteriori* state estimate and its covariance at sample time $t_{k-m(k)}$. They are computed from the linearized forward filtering pass during the last Gauss-Newton iteration of the smoothing problem that ends at sample k. The details of this computation are in Ref. 18.

This algorithm was applied to the difficult problem of a tumbling micro-satellite in a circular earth orbit with altitude 823 km and inclination 82 deg. The satellite uses only magnetometer measurements sampled once every 10 seconds to estimate its attitude and angular rate, which has the initial value of 2 deg/sec. Eulerian dynamics were modelled with gravity-gradient torques and white-noise disturbance torques. The filter estimates the components of the spacecraft inertia tensor, which is not diagonal, along with the satellite attitude and rates. Since the overall scale of the inertia tensor is unobservable, the scale was fixed by assigning a very small *a priori* variance to one of the elements.

Three cases were simulated. All had the same initial estimate of the inertia tensor elements with errors as large as 4.4% of the maximum principal inertia, which resulted in errors in the principal axis orientations of up to 44 deg. The first simulation had moderate initial errors of 0.34 deg in attitude and in 9.1% in angular rate. The second simulation tested the filter's ability to converge from large initial errors by starting with a 180 deg attitude error and a 77% angular rate error. The third simulation was an even more severe test, starting with a rate 9 times the true attitude rate, which causes a 180 deg increase in the attitude error during the filter's 10 second sampling period.

The BSEKF was compared with an EKF and an UF.[30] In the case with moderate initial errors, the BSEKF was tested with m_{target} values of 10, 20, and 30. The best results for this case were returned by the UF and the BSEKF with $m_{\text{target}} = 30$, which both converged to attitude errors of less than 2 deg after about 1 hour of operation with the BSEKF being slightly more accurate. The BSEKF with $m_{\text{target}} = 20$ was about 41% less accurate, and the BSEKF with $m_{\text{target}} = 10$ was about 850% less accurate. The performance of the EKF fell between that of the BSEKF with $m_{\text{target}} = 10$ and with $m_{\text{target}} = 20$. The accuracy of the moment-of-inertia estimation for the different filters followed the same pattern. The poorer performance of the BSEKF with $m_{\text{target}} = 10$ compared with the EKF is unexplained.

Based on these results, only $m_{\text{target}} = 30$ was used in the more stressing simulations, and the BSEKF performed as well in these cases as in first case. The EKF had attitude errors of 20 deg even after 2 hours in both stressing cases. The performance of the UF was between the EKF and the BSEKF in the second case, but it failed to converge at all in the third case. This is believed to be a result of a wrap-around discontinuity of the sigma points for the attitude representation used in the particular UF.

The BSEKF's performance improvements are obtained at the expense of between 165 and 175 times as much processing as the EKF. However, the computation time is only about 18-19 times that of the UF for this problem. The improved performance shows the potential of the BSEKF for nonlinear estimation.

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 RADCAL satellite using GPS measurements.[21]

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

$$\tilde{\omega} = \omega + \beta + B_1 \mathbf{w}_d \tag{35a}$$

$$\dot{\beta} = B_2 \mathbf{w}_d \tag{35b}$$

where $\mathbf{w}_d$ is a square-integrable six-component deterministic dynamic disturbance vector, and B_1 and B_2 are known matrices. It is assumed that $B_1 B_2^T = 0_{3\times3}$, although this assumption is not necessary. The kinematics can be written with $\mathbf{x} = [\mathbf{q}^T \ \beta^T]^T$ as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + G(\mathbf{x})\mathbf{w}_d \tag{36}$$

where

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \frac{1}{2}\Xi(\mathbf{q})(\tilde{\omega} - \beta) \\ 0_{3\times1} \end{bmatrix} \quad \text{and} \quad G(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2}\Xi(\mathbf{q})B_1 \\ B_2 \end{bmatrix} \tag{37}$$

Continuous measurements are modelled as

$$\tilde{\mathbf{y}} = \mathbf{h}(\mathbf{x}) + D(\mathbf{x})\mathbf{w}_y \tag{38}$$

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

$$\dot{\hat{\mathbf{x}}} = \mathbf{f}(\hat{\mathbf{x}}) + K(\hat{\mathbf{x}}, t)[\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}})] \tag{39}$$

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

$$\mathbf{z} = \zeta(\mathbf{x}) - Z^T(\mathbf{x})\hat{\mathbf{x}} \tag{40}$$

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 $K(\hat{\mathbf{x}}, t)$ for which

$$\int_0^T \|\mathbf{z}(\tau)\|^2 \, d\tau \le \gamma^2 \left\{ N(\mathbf{x}_0, \hat{\mathbf{x}}_0) + \int_0^T \left[\|\mathbf{w}_d(\tau)\|^2 + \|\mathbf{w}_y(\tau)\|^2 \right] \, d\tau \right\} \tag{41}$$

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

$$J \equiv \partial V / \partial t + (\partial V / \partial \mathbf{x}) \, \mathbf{f}(\mathbf{x}) + (\partial V / \partial \hat{\mathbf{x}}) \, \{ \mathbf{f}(\hat{\mathbf{x}}) + K(\hat{\mathbf{x}}, t) \, [\mathbf{h}(\mathbf{x}) - \mathbf{h}(\hat{\mathbf{x}})] \} + \mathbf{z}^T \mathbf{z}$$
$$+ \frac{1}{4} \gamma^{-2} \left[(\partial V / \partial \mathbf{x}) \, G(\mathbf{x}) G^T(\mathbf{x}) \, (\partial V / \partial \mathbf{x})^T + (\partial V / \partial \hat{\mathbf{x}}) \, K(\hat{\mathbf{x}}, t) R(\mathbf{x}) K^T(\hat{\mathbf{x}}, t) \, (\partial V / \partial \hat{\mathbf{x}})^T \right] \le 0 \qquad (42)$$

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

For the attitude estimation case, we take $\boldsymbol{\zeta}(\mathbf{x}) = [0_{1 \times 3} \quad \boldsymbol{\beta}^T]^T$ and

$$Z(\mathbf{x}) = \begin{bmatrix} \Xi(\mathbf{q}) & 0_{4 \times 3} \\ 0_{3 \times 3} & I_{3 \times 3} \end{bmatrix} \qquad (43)$$

so that

$$\mathbf{z} = \begin{bmatrix} -\Xi^T(\mathbf{q}) \, \hat{\mathbf{q}} \\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \end{bmatrix} \qquad (44)$$

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

A significant amount of algebra using some small-error approximations shows that Eq. (42) is satisfied for

$$V(\mathbf{x}, \hat{\mathbf{x}}, t) = \gamma^2 \mathbf{z}^T P^{-1} \mathbf{z} \qquad (45)$$

with the gain matrix

$$K(\hat{\mathbf{x}}, t) = Z(\hat{\mathbf{x}}) P \begin{bmatrix} H^T(\hat{\mathbf{q}}) \\ 0_{3 \times m} \end{bmatrix} R^{-1} \qquad (46)$$

where $H(\hat{\mathbf{q}})$ 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

$$-\dot{P} + F(\hat{\mathbf{x}}) P + P F^T(\hat{\mathbf{x}}) + P \begin{bmatrix} \gamma^{-2} I_{3 \times 3} - H^T(\hat{\mathbf{q}}) R^{-1} H(\hat{\mathbf{q}}) & 0_{3 \times 3} \\ 0_{3 \times 3} & \gamma^{-2} I_{3 \times 3} \end{bmatrix} P + \begin{bmatrix} \frac{1}{4} Q_1 & 0_{3 \times 3} \\ 0_{3 \times 3} & Q_2 \end{bmatrix} \le 0 \qquad (47)$$

with

$$F(\hat{\mathbf{x}}) = - \begin{bmatrix} [(\tilde{\boldsymbol{\omega}} - \hat{\boldsymbol{\beta}}) \times] & \frac{1}{2} I_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} \end{bmatrix} \qquad (48a)$$

$$Q_1 = B_1 B_1^T \quad \text{and} \quad Q_2 = B_2 B_2^T \qquad (48b)$$

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. (41) and (47), 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. (47) goes to the usual Riccati equation of the Kalman filter, and the deterministic estimator becomes the continuous measurement limit of the MEKF.[6]

G. Interlaced Kalman Filter

The basic idea of the interlaced Kalman filter[27–29] (IKF) is to replace a nonlinear model by two pseudo-linear models and to estimate the state by using two interlaced linear Kalman filters running simultaneously. The application of interfaced filters to spacecraft has been to smooth noisy gyro measurements[27, 28] or to estimate attitude rates in the absence of gyro data.[29]

These applications begin with the dynamics equation for a rigid spacecraft equipped with reaction wheels[92]

$$\mathcal{J} \dot{\boldsymbol{\omega}} + \dot{\mathbf{h}} + [\boldsymbol{\omega} \times] (\mathcal{J} \boldsymbol{\omega} + \mathbf{h}) = \mathbf{T} \qquad (49)$$

where $\mathcal{J}$ is the moment-of-inertia tensor, $\mathbf{h}$ is the reaction wheel angular momentum, and $\mathbf{T}$ is the external torque, which includes disturbances and magnetic and propulsive control torques, but not reaction wheel

torques. It is most convenient to analyze the dynamics in the principal axis frame, in which $\mathcal{J}$ is diagonal and

$$\dot{\omega} = \mathcal{J}^{-1}[\mathbf{h}\times]\omega + G_\omega \left[\omega_2\omega_3 \ \ \omega_1\omega_3 \ \ \omega_1\omega_2\right]^T + \mathcal{J}^{-1}(\mathbf{T} - \dot{\omega}) \tag{50}$$

with

$$G_\omega = \begin{bmatrix} \mathcal{J}_{11}^{-1}(\mathcal{J}_{22} - \mathcal{J}_{33}) & 0 & 0 \\ 0 & \mathcal{J}_{22}^{-1}(\mathcal{J}_{33} - \mathcal{J}_{11}) & 0 \\ 0 & 0 & \mathcal{J}_{33}^{-1}(\mathcal{J}_{11} - \mathcal{J}_{22}) \end{bmatrix} \tag{51}$$

This is a nonlinear equation in ω, but the interlaced filter converts it into a linear equation by estimating $\chi \equiv \left[\omega_2\omega_3 \ \ \omega_1\omega_3 \ \ \omega_1\omega_2\right]^T$ in a second Kalman filter, and writing

$$\dot{\omega} = \mathcal{J}^{-1}[\mathbf{h}\times]\omega + G_\omega\hat{\chi} + \mathcal{J}^{-1}(\mathbf{T} - \dot{\omega}) \tag{52}$$

where $\hat{\chi}$ is the estimate obtained in the second filter. The dynamic equation of this filter is

$$\dot{\chi} = G_\chi\hat{\omega} \tag{53}$$

where $\hat{\omega}$ is the estimate obtained by the first filter and

$$G_\chi = \begin{bmatrix} 0 & \dot{\omega}_3 & \dot{\omega}_2 \\ \dot{\omega}_3 & 0 & \dot{\omega}_1 \\ \dot{\omega}_2 & \dot{\omega}_1 & 0 \end{bmatrix} \tag{54}$$

is given as a nonlinear function of $\hat{\omega}$ by Eq. (50). The IKF comprises two linear Kalman filters running simultaneously, one with dynamics given by Eq. (52) to estimate the attitude rate vector ω, and the other with dynamics given by Eq. (53) to estimate the products of the components contained in the state vector χ.

Reference 27 treated the case of noisy gyro measurements, for which the measurement inputs to the two interlaced filters are trivially obtained from the gyro outputs. Both the IKF and the EKF were very effective in filtering out the measurement noise for this problem, but the attitude rate estimation errors of the interlaced filter were about 25% less on the pitch axis and about 33% less on the roll and yaw axes. The interlaced filter ignores correlations in the process noise and the measurement errors in the two filters, but it appears that this approximation is less problematical than the linearizations in the EKF.

Reference 29 treated the more challenging gyroless case, for which rate estimates must be obtained by differentiating vector measurements. Differentiating the noiseless equation $\mathbf{b}_i = A\mathbf{r}_i$ gives the well-known equation

$$\dot{\mathbf{b}}_i = -\omega \times \mathbf{b}_i + A\dot{\mathbf{r}}_i \tag{55}$$

This can be written as a measurement equation for ω

$$\mathbf{y}_i = C_i\omega \tag{56}$$

with

$$\mathbf{y}_i \equiv \dot{\mathbf{b}}_i - A\dot{\mathbf{r}}_i \quad \text{and} \quad C_i \equiv [\mathbf{b}_i\times] \tag{57}$$

If two vectors are measured simultaneously then a pointwise estimate of ω can be computed by

$$\tilde{\omega} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \tag{58}$$

where the dagger denotes the Moore-Penrose pseudo-inverse. Note that $\tilde{\omega}$ is a "measurement" to be used in the filter. The generalization to more than two vectors is obvious. Producing inputs to both of the interlaced filters from $\tilde{\omega}$ can smooth this estimate in the same way that Ref. 27 smoothes gyro measurements.

Equation (56) can also be used as a measurement for the linear filter for ω. There is no other option if only one vector at a time is available. Complications arising from the fact that C_i is rank-deficient can be avoided by deleting one component of and the corresponding row of C_i. There are two options for obtaining measurements for the χ filter. The simplest is to simply use products of the angular rate estimates from the

other interlaced filter. Independent measurements can be derived at the expense of more computations by using second time derivatives of the vector measurements.

The interlaced filter employing first and second derivatives of magnetometer and sun sensor measurements was tested using flight data from the Rossi X-Ray Timing Explorer (RXTE) spacecraft. Gyro data were available to assess the filter performance, but were not used in the filter. The data span analyzed included a 90 deg maneuver around the z axis at a rate of 0.125 deg/sec. Magnetometer data were available for the entire data span, but sun sensor data became available only 204 seconds after the start of the maneuver. The maximum rate estimation errors were about 0.03 deg/sec in all three axes with magnetometer data only. When sun sensor data became available, the errors about z axis remained about the same, because the sun vector was very close to this axis, but the errors about the x and y axes were reduced by an order of magnitude. Adding the cross product of the sun vector and the magnetic field vector as a third measurement reduced the z axis errors to 0.004 deg/sec when the sun was in view, and adding the cross-product of the magnetic field vector and its time derivative when the sun was unavailable decreased the x and y errors to about 0.005 deg/sec during these periods. The details are in Reference 29, which did not include a comparison with the EKF for this problem.

Two other approaches that do use an EKF to estimate angular rates are presented in Ref. 93. The first uses differentiated quaternion measurements to yield a coarse rate measurement, which are then fed into two different estimators. The other approach using raw quaternion measurements that are fed directly into two estimators. The nonlinear part of the rotational dynamics equation, shown in Eq. (49), is decomposed to produce a product of an angular-rate dependent matrix and the angular-rate vector itself, which can be used in a pseudo-linear Kalman filter. Results using data from the RXTE spacecraft are also shown in Ref. 93.

IV.   QUEST-Based Methods

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

$$J(A) = \frac{1}{2} \sum_{i=1}^{m} a_i \| \tilde{\mathbf{b}}_i - A\mathbf{r}_i \|^2 \tag{59}$$

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

$$J(A) = \lambda_0 - \text{trace}(AB^T) \tag{60}$$

with

$$\lambda_0 \equiv \sum_{i=1}^{m} a_i \tag{61}$$

and

$$B \equiv \sum_{i=1}^{m} a_i \tilde{\mathbf{b}}_i \mathbf{r}_i^T \tag{62}$$

makes it is 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 pointed out 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.[96]

There are many algorithms for solving this problem,[97] of which the most useful are Davenport's q method[98] and QUEST.[23] Davenport parameterized the attitude matrix by a unit quaternion, as shown by Eq. (4), giving

$$\text{trace}(AB^T) = \mathbf{q}^T K \mathbf{q} \tag{63}$$

where K is the symmetric traceless matrix

$$K \equiv \begin{bmatrix} B + B^T - \text{trace}(B)I_{3\times3} & \sum_{i=1}^{m} a_i \tilde{\mathbf{b}}_i \times \mathbf{r}_i \\ \left(\sum_{i=1}^{m} a_i \tilde{\mathbf{b}}_i \times \mathbf{r}_i \right)^T & \text{trace}(B) \end{bmatrix} = -\sum_{i=1}^{m} a_i \Omega(\tilde{\mathbf{b}}_i)\Gamma(\mathbf{r}_i) \tag{64}$$

The relation on the right-hand-side of Eq. (64) 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 properties.[99] The optimal attitude is represented by the quaternion maximizing the right-hand-side of Eq. (63), subject to the unit constraint $\|\mathbf{q}\| = 1$. It is not difficult to see that the optimal quaternion is equal to the normalized eigenvector of K with the largest eigenvalue, i.e. the solution of

$$K\mathbf{q}_{\text{opt}} = \lambda_{\max}\mathbf{q}_{\text{opt}} \tag{65}$$

With Eqs. (60) and (63), this gives the optimized loss function as

$$J(A_{\text{opt}}) = \lambda_0 - \lambda_{\max} \tag{66}$$

The very efficient QUEST algorithm is based on Shuster's observation that $\lambda_{\max}$ can be easily obtained by a Newton-Raphson iteration starting from λ_0 as the initial estimate, since Eq. (66) 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 many times over in practical applications.

A. Filter QUEST

Since a filtering algorithm is usually preferred when observations are obtained over a range of times, Shuster proposed the filter QUEST algorithm,[24] based on propagating and updating B:

$$B(t_k) = \mu\Phi_{3\times 3}(t_k,\, t_{k-1})B(t_{k-1}) + \sum_{i=1}^{m_k} a_i\tilde{\mathbf{b}}_i\mathbf{r}_i^T \tag{67}$$

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 QUEST or REQUEST,[25] propagates and updates Davenport's K matrix by

$$K(t_k) = \mu\Phi_{4\times 4}(t_k,\, t_{k-1})K(t_{k-1})\Phi_{4\times 4}^T(t_k,\, t_{k-1}) + \sum_{i=1}^{m_k} a_iK_i \tag{68}$$

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:

$$K_i = \begin{bmatrix} \tilde{\mathbf{b}}_i\mathbf{r}_i^T + \mathbf{r}_i\tilde{\mathbf{b}}_i^T - (\tilde{\mathbf{b}}_i^T\mathbf{r}_i)I_{3\times 3} & (\tilde{\mathbf{b}}_i \times \mathbf{r}_i) \\ \\ (\tilde{\mathbf{b}}_i \times \mathbf{r}_i)^T & \tilde{\mathbf{b}}_i^T\mathbf{r}_i \end{bmatrix} \tag{69}$$

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.[100]

B. Extended QUEST

Extended QUEST[26] is an algorithm that solves for the attitude along with additional parameters. This is accomplished by finding the attitude quaternion $\mathbf{q}_k$ and the vector of auxiliary filter states $\mathbf{x}_k$, along with $\mathbf{q}_{k-1}$, $\mathbf{x}_{k-1}$, and the process noise vector $\mathbf{w}_{k-1}$ that minimize the loss function

$$J = \frac{1}{2}\sum_{i=1}^{m_k}\sigma_i^{-2}\|\tilde{\mathbf{b}}_i - A(\mathbf{q}_k)\mathbf{r}_i\|^2 + \frac{1}{2}\|R_{\mathbf{w}\mathbf{w}(k-1)}\mathbf{w}_{k-1}\|^2 + \frac{1}{2}\|R_{\mathbf{q}\mathbf{q}(k-1)}(\mathbf{q}_{k-1} - \hat{\mathbf{q}}_{k-1})\|^2$$
$$+ \frac{1}{2}\|R_{\mathbf{x}\mathbf{q}(k-1)}(\mathbf{q}_{k-1} - \hat{\mathbf{q}}_{k-1}) + R_{\mathbf{x}\mathbf{x}(k-1)}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})\|^2 \tag{70}$$

subject to the attitude dynamics equation

$$\mathbf{q}_k = \Phi(t_k,\, t_{k-1};\, \mathbf{q}_{k-1},\, \mathbf{x}_{k-1},\, \mathbf{w}_{k-1})\mathbf{q}_{k-1} \tag{71}$$

the transition equation for the auxiliary filter states,

$$\mathbf{x}_k = \mathbf{f}_\mathbf{x}(t_k, t_{k-1}; \mathbf{q}_{k-1}, \mathbf{x}_{k-1}, \mathbf{w}_{k-1}) \tag{72}$$

and the norm constraint $||\mathbf{q}_k|| = 1$. The vectors $\hat{\mathbf{q}}_{k-1}$ and $\hat{\mathbf{x}}_{k-1}$ are the *a posteriori* (or best) estimates of $\mathbf{q}$ and $\mathbf{x}$ 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[91] 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:

$$\tilde{\mathbf{q}}_k = \Phi(t_k, t_{k-1}; \hat{\mathbf{q}}_{k-1}, \hat{\mathbf{x}}_{k-1}, \mathbf{0})\hat{\mathbf{q}}_{k-1} \tag{73}$$

and

$$\tilde{\mathbf{x}}_k = \mathbf{f}_\mathbf{x}(t_k, t_{k-1}; \hat{\mathbf{q}}_{k-1}, \hat{\mathbf{x}}_{k-1}, \mathbf{0}) \tag{74}$$

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

$$J(\mathbf{q}_k, \mathbf{x}_k) = \frac{1}{2}\sum_{i=1}^{m_k}\sigma_i^{-2}||\tilde{\mathbf{b}}_i - A(\mathbf{q}_k)\mathbf{r}_i||^2 + \frac{1}{2}||\tilde{R}_{\mathbf{qq}(k)}(\mathbf{q}_k - \tilde{\mathbf{q}}_k)||^2 + \frac{1}{2}||\tilde{R}_{\mathbf{xq}(k)}(\mathbf{q}_k - \tilde{\mathbf{q}}_k) + \tilde{R}_{\mathbf{xx}(k)}(\mathbf{x}_k - \tilde{\mathbf{x}}_k)||^2 \tag{75}$$

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. Thus Eq. (75) is an approximation to Eq. (70) if the dynamics are nonlinear, but is exact for linear dynamics.

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

$$(\mathbf{x}_k)_{\mathrm{opt}} = \tilde{\mathbf{x}}_k - \tilde{R}_{\mathbf{xx}(k)}^{-1}\tilde{R}_{\mathbf{xq}(k)}(\mathbf{q}_k - \tilde{\mathbf{q}}_k) \tag{76}$$

Substituting this in Eq. (75) and using Eqs. (60), (62) and (69) gives

$$J(\mathbf{q}_k, \hat{\mathbf{x}}_k) = -\mathbf{q}_k^T\left(\sum_{i=1}^{m_k}\sigma_i^{-2}K_i\right)\mathbf{q}_k + \frac{1}{2}\left[\tilde{R}_{\mathbf{qq}(k)}(\mathbf{q}_k - \tilde{\mathbf{q}}_k)\right]^T\left[\tilde{R}_{\mathbf{qq}(k)}(\mathbf{q}_k - \tilde{\mathbf{q}}_k)\right] \tag{77}$$

Minimizing this loss function gives the best estimate $\hat{\mathbf{q}}_k$. The minimization differs from Wahba's problem because of the linear terms in $\mathbf{q}_k$. Substituting $\hat{\mathbf{q}}_k$ into Eq. (76) gives

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k - \tilde{R}_{\mathbf{xx}(k)}^{-1}\tilde{R}_{\mathbf{xq}(k)}(\hat{\mathbf{q}}_k - \tilde{\mathbf{q}}_k) \tag{78}$$

The remaining step is to express the optimized loss function in the form of Eq. (70) to prepare for the next step in the recursion. The details are in Ref. 26.

This algorithm was tested in a simulation of a rigid spacecraft in near-earth orbit equipped with a star tracker and a rate gyro. Eulerian dynamics were modelled with gravity-gradient torques and white-noise disturbance torques. In one scenario, the spacecraft was spinning and nutating with spin period of 50 sec and nutation period of 504 sec. The other simulation was of a nadir-pointing spacecraft undergoing libration at roughly orbit frequency. The extended QUEST algorithm was shown to perform significantly better than the EKF, at about 2.6 times the computational expense.

V. Two-Step Attitude Estimator

The two-step optimal estimator was proposed in 1995 as an alternative to the standard EKF.[101] An implementation of the two-step optimal estimation for recursive spacecraft attitude estimation is presented in Ref. 102. Simulations show dramatic improvement of the two-step attitude estimator over the traditional EKF and the extended QUEST.[102]

The general cost function of the two-step optimal estimator is given by

$$J_k = \frac{1}{2}\sum_{i=0}^{k-1}\left\{[\tilde{\mathbf{y}}_{i+1} - \mathbf{h}_{i+1}(\mathbf{x}_{i+1})]^T R_{i+1}^{-1}[\tilde{\mathbf{y}}_{i+1} - \mathbf{h}_{i+1}(\mathbf{x}_{i+1})] + \mathbf{w}_i^T Q_i^{-1}\mathbf{w}_i\right\}$$
$$+ \frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T P_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) \tag{79}$$

subject to the dynamics equation

$$\mathbf{x}_{i+1} = \mathbf{f}_i(\mathbf{x}_i, \mathbf{w}_i) \quad \text{for} \quad i = 0, 1, \ldots, k-1 \tag{80}$$

The process noise covariance is Q_i, $\tilde{\mathbf{y}}_{i+1}$ is the measurement with covariance R_{i+1}, $\mathbf{h}_{i+1}(\mathbf{x}_{i+1})$ is the nonlinear measurement model, and $\hat{\mathbf{x}}_0$ is the *a priori* estimate of the state with covariance P_0. The state estimate $\hat{\mathbf{x}}_k$ 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 $\mathcal{Y} = \mathcal{F}(\mathbf{x})$ in which the nonlinear measurement model is linear, i.e. $\tilde{\mathbf{y}} = \mathbf{h}(\mathbf{x}) + \mathbf{v} = \mathcal{H}\mathcal{Y} + \mathbf{v}$, where $\mathbf{v}$ is the measurement noise, so that a linear measurement update of the first-step state can be applied. The first-step state $\mathcal{Y}$ relates the desired or second-step state $\mathbf{x}$ and the measurement $\tilde{\mathbf{y}}$ 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 $\hat{\mathbf{x}}_0$ and P_0, e.g. using a Monte Carlo approach. In general, $(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T P_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)$ 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 $\mathbf{x}$ and $\mathcal{Y}$. The desired state estimate $\hat{\mathbf{x}}_k$ 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

$$J_{\mathbf{x}} = \frac{1}{2} \left[\hat{\mathcal{Y}}_k - \mathcal{F}(\mathbf{x}) \right]^T \mathcal{P}_k^{-1} \left[\hat{\mathcal{Y}}_k - \mathcal{F}(\mathbf{x}) \right] \tag{81}$$

The constraints in $\hat{\mathbf{x}}_k$ may also be included by use of the Lagrangian multiplier method. A numerical least-squares algorithm such as the Gaussian-Newton or Levenberg-Marquardt method is used to solve the second-step minimization problem. Note that in order 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 is no dynamics, i.e. $\mathbf{x}_{k+1} = \mathbf{x}_k$, or initial guess, Eq. (79) reduces to

$$J_k = \frac{1}{2} \sum_{i=1}^{k} \left[\tilde{\mathbf{y}}_i - \mathbf{h}_i(\mathbf{x}) \right]^T R_i^{-1} \left[\tilde{\mathbf{y}}_i - \mathbf{h}_i(\mathbf{x}) \right] \tag{82}$$

The two-step attitude estimator for this special case amounts to 1) formulating the cost function in terms of $\tilde{\mathbf{y}}_i$ 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 $\hat{\mathbf{x}}_k$ 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.

The desired or second-step state of the two-step attitude estimator 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

$$\mathcal{Y} = [A_{11}, A_{12}, A_{13}, A_{21}, ..., A_{33}]^T \tag{83}$$

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. (83) to a unit vector observation is given by

$$\mathcal{H} = \begin{bmatrix} \mathbf{r}^T & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} \\ \mathbf{0}_{1 \times 3} & \mathbf{r}^T & \mathbf{0}_{1 \times 3} \\ \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{r}^T \end{bmatrix} \tag{84}$$

where $\mathbf{r}$ 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 the

American Institute of Aeronautics and Astronautics

six constraints in the attitude matrix have not been accounted for.[102] 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 Ref. 102.

The dynamic propagation equation of the attitude part of the first-step state, $\mathcal{Y}$ in Eq. (83), is given by

$$\dot{\mathcal{Y}}(t) = \tilde{\Omega}(\omega)\mathcal{Y}(t) \tag{85}$$

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

$$\tilde{\Omega}(\omega) = \begin{bmatrix} 0_{3\times3} & \omega_3 I_{3\times3} & -\omega_2 I_{3\times3} \\ -\omega_3 I_{3\times3} & 0_{3\times3} & \omega_1 I_{3\times3} \\ \omega_2 I_{3\times3} & -\omega_1 I_{3\times3} & 0_{3\times3} \end{bmatrix} \tag{86}$$

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 in order 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.[102]

When the angular rate ω in Eq. (85) is perfect known, only a linear Kalman filter needs to be implemented for the first-step state, i.e. $\mathcal{Y}$ in Eq. (83). When the angular rate is obtained from noisy gyro measurements and both the attitude and the gyro biases are to be estimated, the augmented first-step state dynamics becomes nonlinear and a nonlinear time-update scheme should be implemented for the first-step state because of the state-dependent process noise and the product terms of the entries of the attitude matrix and the gyro biases in the dynamic propagation equation.

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 Ref. 87, although the gyro measurement and vector measurement models used therein are slightly different. In Ref. 87, 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 Ref. 87 is not always assured when they are initialized with the identity matrix. Whether the two-step attitude estimator has guaranteed convergence in the same scenario is unclear.

VI.   Unscented Filtering

The Unscented filter[30,31] 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

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, k) + \mathbf{w}_k \tag{87a}$$

$$\tilde{\mathbf{y}}_k = \mathbf{h}(\mathbf{x}_k, k) + \mathbf{v}_k \tag{87b}$$

Note that a continuous-time model can always be written using Eq. (87a) through an appropriate numerical integration scheme. It is again assumed that $\mathbf{w}_k$ and $\mathbf{v}_k$ are zero-mean Gaussian noise processes with covariances given by Q_k and R_k, respectively. The Kalman gain and updated covariance are rewritten in the form given by[103]

$$K_k = P_k^{xy}(P_k^{vv})^{-1} \tag{88a}$$

$$P_k^+ = P_k^- - K_k P_k^{vv} K_k^T \tag{88b}$$

where P_k^+ and P_k^- are the updated and propagated covariances, respectively, P_k^{vv} is the covariance of the innovations process $\tilde{\mathbf{y}}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-, k)$ with $\hat{\mathbf{x}}_k^-$ denoting the propagated state, and P_k^{xy} is the cross-correlation matrix between $\hat{\mathbf{x}}_k^-$ and $\mathbf{h}(\hat{\mathbf{x}}_k^-, k)$. The general formulation for the propagation equations and computation of the gain matrix is given by computing the following *sigma points*:

$$\boldsymbol{\sigma}_k \leftarrow 2n \text{ columns from } \pm\gamma\sqrt{P_k^+} \tag{89a}$$

$$\chi_k(0) = \hat{\mathbf{x}}_k^+ \tag{89b}$$

$$\chi_k(i) = \boldsymbol{\sigma}_k(i) + \hat{\mathbf{x}}_k^+ \tag{89c}$$

where γ is a design parameter and $\sqrt{P_k^+}$ denotes the matrix square root of P_k^+.[95] Due to the symmetric nature of this set, its odd central moments are zero, 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

$$\chi_{k+1}(i) = \mathbf{f}(\chi_k(i), k) \quad \text{for} \quad i = 0, 1, \dots, 2n \tag{90}$$

The predicted mean at times t_{k+1} for the state estimate is calculated using a weighted sum of the points $\chi_{k+1}^x(i)$, which is given by

$$\hat{\mathbf{x}}_{k+1}^- = \sum_{i=0}^{2n} W_i^{\mathrm{mean}} \chi_{k+1}(i) \tag{91}$$

where W_i^{mean} is a weighting parameter. The output covariance and cross-correlation are computed by

$$P_{k+1}^{yy} = \sum_{i=0}^{2n} W_i^{\mathrm{cov}} \left[\boldsymbol{\gamma}_{k+1}(i) - \hat{\mathbf{y}}_{k+1}^-\right] \left[\boldsymbol{\gamma}_{k+1}(i) - \hat{\mathbf{y}}_{k+1}^-\right]^T \tag{92a}$$

$$P_{k+1}^{xy} = \sum_{i=0}^{2n} W_i^{\mathrm{cov}} \left[\chi_{k+1}(i) - \hat{\mathbf{x}}_{k+1}^-\right] \left[\boldsymbol{\gamma}_{k+1}(i) - \hat{\mathbf{y}}_{k+1}^-\right]^T \tag{92b}$$

where W_i^{cov} is another weighting parameter and

$$\boldsymbol{\gamma}_{k+1}(i) = \mathbf{h}(\chi_{k+1}(i), k+1) \tag{93}$$

Since the measurement noise appears linearly in Eq. (87b), then $P_{k+1}^{vv} = P_{k+1}^{yy} + R_{k+1}$. Methods to handle process noise for the computation of the propagated covariance and to handle nonlinearly appearing measurement noise are discussed in Ref. 31.

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. (91). To overcome this problem an unconstrained three-component vector is used, based on the generalized Rodrigues parameters[77] (GRPs) to represent an attitude error quaternion.[33] The algorithm is called the UnScented QUaternion Estimator, or USQUE. The state vector includes the attitude error and bias vectors:

$$\chi_k(0) = \hat{\mathbf{x}}_k^+ \equiv \begin{bmatrix} \delta\hat{\mathbf{p}}_k^+ \\ \hat{\boldsymbol{\beta}}_k^+ \end{bmatrix} \tag{94}$$

where the estimated error-GRP $\delta\hat{\mathbf{p}}_k$ 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 $\delta\mathbf{q}_k^+(i)$, through a simple transformation from GRPs to quaternions. Then, the following quaternions are computed:

$$\hat{\mathbf{q}}_k^+(0) = \hat{\mathbf{q}}_k^+ \tag{95a}$$

$$\hat{\mathbf{q}}_k^+(i) = \delta\mathbf{q}_k^+(i) \otimes \hat{\mathbf{q}}_k^+, \quad i = 1, 2, \dots, 12 \tag{95b}$$

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, so we might expect the covariance to 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, since

American Institute of Aeronautics and Astronautics

the update is zero-mean, the mean rotation caused by the reset is actually zero, so the covariance is in fact not affected by the reset. The quaternions in Eq. (95) are propagated using Eq. (7) with the estimated angular rate. The propagated error-quaternions are then computed using

$$\delta \mathbf{q}_{k+1}^-(i) = \hat{\mathbf{q}}_{k+1}^-(i) \otimes \left[\hat{\mathbf{q}}_{k+1}^-(0)\right]^{-1}, \quad i = 0, 1, \ldots, 12 \tag{96}$$

Note that $\delta \mathbf{q}_{k+1}^-(0)$ 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).

A comparison between the standard EKF approach of Ref. 6 and the USQUE algorithm has been shown in Ref. 33. The simulation involves determining the attitude of a low-Earth pointing spacecraft from magnetometer and gyro measurements only. With good initial attitude estimates, both USQUE and the EKF attitude errors agree to within 1 μrad. This indicates that no advantages to using USQUE with small errors can be seen. Another simulation involved large attitude errors of $-50°$, $50°$ and $160°$ for each axis, respectively, with initial bias estimates set to zero. The EKF takes almost 8 hours to converge to a value below $0.1°$, while USQUE converges to this value in under 30 min. Also, the EKF attitude errors do not converge to within their respective 3σ error bounds until well after 8 hours. However, the USQUE attitude errors do converge to within their respective 3σ well within 8 hours. Another case is shown where the EKF never converges, while USQUE does.

VII.   Particle Filters

There is no such thing as "the PF," just as there is no such thing as "the EKF."[104] 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 pseudo-random 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 Refs. 34, 35 and 105.

A PF allows for a more general discrete-time state-space model than the UF, given by

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \tag{97a}$$

$$\tilde{\mathbf{y}}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k) \tag{97b}$$

The process noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are assumed to be white noise processes. The distributions of $\mathbf{x}_0$, $\mathbf{w}_k$, and $\mathbf{v}_k$, denoted by $p(\mathbf{x}_0)$, $p(\mathbf{w}_k)$, and $p(\mathbf{v}_k)$, respectively, are assumed to be known and mutually independent. No Gaussian assumptions about the noises are required. The probabilities $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and $p(\mathbf{y}_k|\mathbf{x}_k)$ can be derived from the above model and are assumed to be available for sampling and evaluation.

The empirical, discrete approximation of the posterior distribution $p(\mathbf{x}_k|\mathbf{Y}_k)$ with N weighted particles $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ is given by

$$P_N(d\mathbf{x}_k|\mathbf{Y}_k) \approx \sum_{i=1}^N w_k^{(i)} \delta_{\mathbf{x}_k^{(i)}}(d\mathbf{x}_k) \tag{98}$$

where $\mathbf{x}_k^{(i)}$ are the particles drawn from the importance function or proposal distribution, $w_k^{(i)}$ are the normalized importance weights, satisfying $\sum_{i=1}^N w_k^{(i)} = 1$, and $\delta_{\mathbf{x}_k^{(i)}}(d\mathbf{x}_k)$ denotes the Dirac-delta mass located in $\mathbf{x}_k^{(i)}$. We use $\mathbf{X}_k$ and $\mathbf{Y}_k$ to denote the state trajectory $\{\mathbf{x}_j\}_{j=0}^k$ and measurement history $\{\mathbf{y}_j\}_{j=1}^k$, respectively. The expectation of a known function $\mathbf{f}(\mathbf{x}_k)$ with respect to $p(\mathbf{x}_k|\mathbf{Y}_k)$ is then approximated by

$$\int \mathbf{f}(\mathbf{x}_k) p(\mathbf{x}_k) d\mathbf{x}_k \approx \sum_{i=1}^N w_k^{(i)} \mathbf{f}(\mathbf{x}_k^{(i)}) \tag{99}$$

For example, the approximation to the arithmetic mean of $\mathbf{x}_k$ is $\sum_{i=1}^N w^{(i)} \mathbf{x}^{(i)}$. Crisan and Doucet showed that the upper bound on the variance of the estimation error has the form $c\,\mathcal{O}(N^{-1})$, with c a constant,[105, 106] whereas Daum argued that c in the upper bound depends heavily on the state vector dimension.[104, 105]

A PF updates the particle representation $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ in a recursive manner. A cycle of a generic PF includes[105]

- Sequential Importance Sampling

    - For $i = 1, \dots N$, sample $\mathbf{x}_{k+1}^{(i)}$ from the importance function $q(\mathbf{x}_{k+1}|\mathbf{X}_k^{(i)}, \mathbf{Y}_{k+1})$
    - For $i = 1, \dots N$, evaluate and normalize the importance weights

$$w_{k+1}^{(i)} \propto w_k^{(i)} \frac{p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}^{(i)})p(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)})}{q(\mathbf{x}_{k+1}^{(i)}|\mathbf{X}_k^{(i)}, \mathbf{Y}_{k+1})} \tag{100}$$

- Reampling: Multiple/Discard particles $\{\mathbf{x}_{k+1}^{(i)}\}_{i=1}^N$ with respect to high/low importance weights $w_{k+1}^{(i)}$ to obtain N new particles $\{\mathbf{x}_{k+1}^{(i)}\}_{i=1}^N$ with equal weights.

A popular suboptimal choice of the importance function is $q(\mathbf{x}_{k+1}|\mathbf{X}_k^{(i)}, \mathbf{Y}_{k+1}) = p(\mathbf{x}_{k+1}|\mathbf{x}_k^{(i)})$. The PF with this importance function is known as the bootstrap filter (BF). Sampling $\mathbf{x}_{k+1}^{(i)}$ from $p(\mathbf{x}_{k+1}|\mathbf{x}_k^{(i)})$ is equivalent to the dynamic propagation of $\mathbf{x}_k^{(i)}$ to time t_{k+1} in the sampling process. The information contained in the measurement $\mathbf{y}_{k+1}$ is not employed. The corresponding update of the importance weight, Eq. (100), has a simple form, i.e. $w_{k+1}^{(i)} \propto w_k^{(i)} p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}^{(i)})$, which is appealing especially when the evaluation of $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ is difficult. Note that this choice becomes inefficient when the overlap between $p(\mathbf{x}_{k+1}|\mathbf{x}_k^{(i)})$ and $p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1})$ is small. In contrast with this simple importance function, the optimal importance function $p(\mathbf{x}_{k+1}|\mathbf{x}_k^{(i)}, \mathbf{y}_{k+1})$ that minimizes the variance of the importance weight $w_k^{(i)}$ conditional upon $\mathbf{x}_k^{(i)}$ and $\mathbf{y}_{k+1}$ fully incorporates the latest measurement $\mathbf{y}_{k+1}$, 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.[105]

Particle filters are superior to conventional nonlinear filters for difficult applications of nonlinear and non-Gaussian filtering. A frequently-cited illustrative example in the PF literature is repeated here[34]

$$x_{k+1} = \frac{1}{2}x_k + 25\frac{x_k}{1 + x_k^2} + 8\cos[1.2(k+1)] + w_k \tag{101}$$

$$y_k = \frac{x_k^2}{20} + v_k \tag{102}$$

Although the process noise w_k, measurement noise v_k, and the initial error in the above model are independent Gaussian white noises, the posterior distribution of x_k is highly non-Gaussian, and the minimum mean-square error (MMSE) estimate could be misleading. In fact, even if the posterior state distribution is (eventually) Gaussian-like, conventional filters such as the EKF are usually not guaranteed to estimate the mean and covariance correctly. On the other hand, PFs are 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 KF 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 Refs. 36, 37, 107 and 108. 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 Ref. 36, a simple BF was designed to simultaneously estimate the attitude and the gyro biases (or the attitude and the attitude rate in gyro-less applications). Using the uniform attitude distribution as the initial attitude distribution for the case of no initial attitude knowledge and a gradually decreasing measurement variance in the computation of the importance weights, it has good convergence properties. The design of the decreasing variance is heuristic though. This application showed that even a simple BF can be applied to tackle a nontrivial attitude estimation problem.

In Refs. 37 and 107, 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 Ref. 108, which estimates the measurement noise distribution on the fly, along with the estimation of the spacecraft attitude

American Institute of Aeronautics and Astronautics

and gyro biases. The noise distribution estimation scheme is based on an analysis of the filter-generated innovations process.

In Refs. 37,107 and 108 the problem of attitude and gyro bias estimation is divided into two easier problems. The resulting GA-QPF has an interlaced structure. The 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[37]

$$\prod_{j=k_1}^{k_2} p[\mathbf{y}_j | \Phi(\tilde{\omega}_j - \beta)\hat{\mathbf{q}}_{j-1}] \tag{103}$$

where $\Phi(\cdot)$ is the quaternion transition matrix derived from the continuous-time quaternion kinematics and $\hat{\mathbf{q}}_{j-1}$ is obtained from the QPF. In the likelihood function, the gyro bias β is treated as constant over $[t_{k_1}, t_{k_2}]$. The gyro noise model does not need to be incorporated in 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 one degree of freedom. This degree of freedom is used to generate the initial set of particles from the first observation only. By so doing, the QPF starting with 1500 quaternion particles reduces the number to 200 after the first two measurement updates.

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 $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$. 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 ways of computing the attitude estimate have been proposed:

1. As the usual arithmetic mean of three-component attitude representations such as the MRPs,[36] or

2. As the attitude particle with the largest importance weight,[108] or

3. As the minimum of the cost function[108]

$$\sum_{i=1}^N w_k^{(i)} ||A(q_k^{(i)}) - \hat{A}||^2 \tag{104}$$

   subject to $\hat{A}^T \hat{A} = \hat{A}\hat{A}^T = I_{3\times3}$.

The third one is the MMSE estimate in SO(3) and can be computed using the SVD method.[108] The metric in SO(3) associated with the cost function is given by[109]

$$d_F(A_1, A_2) = ||A_1 - A_2|| \tag{105}$$

which is in general not identical with the Riemannian metric in SO(3) corresponding to the length of the shortest geodesic curve, given by[109]

$$d_R(A_1, A_2) = ||\log(A_1^T A_2)|| \tag{106}$$

The third one is superior to the other two but is also more computationally expensive. A problem with the second way 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 way of computing the attitude estimate is also known to be inaccurate. The first choice is not invariant under rotations or attitude parameterization-independent; 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 locally distributed.

American Institute of Aeronautics and Astronautics

VIII. Orthogonal Attitude Filter

The orthogonal filter[40] 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.[110] 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[39] propagates the PDF between measurements and Bayes' formula[1] incorporates measurement information. This approach is related to earlier work by Daum[38] and Lo.[111-113]

Three estimates are known to be identical for a Gaussian PDF defined on $\mathbb{R}^N$: the conditional expectation

$$\bar{\mathbf{x}}_{\mu/v} \equiv \int_{\mathbb{R}^N} \mathbf{x}\, p_{\mathbf{x}_\mu | Y_v}(\mathbf{x})\, d^N\mathbf{x} \tag{107}$$

the MAP estimate

$$\hat{\mathbf{x}}_{\mu/v} \equiv \arg\max_{\mathbf{x} \in \mathbb{R}^N} \left[p_{\mathbf{x}_\mu | Y_v}(\mathbf{x}) \right] \tag{108}$$

and the MMSE estimate

$$\mathbf{x}_{\mu/v}^{\text{MMSE}} \equiv \arg\min_{\mathbf{x}' \in \mathbb{R}^N} \left[\int_{\mathbb{R}^N} \|\mathbf{x} - \mathbf{x}'\|^2\, p_{\mathbf{x}_\mu | Y_v}(\mathbf{x})\, d^N\mathbf{x} \right] \tag{109}$$

It is also 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.

We define the non-Gaussian PDF of the attitude on SO(3) by

$$p_{A_\mu | Y_v}(A) = \exp\left[-J_{A_\mu | Y_v}(A) \right] \tag{110}$$

where the negative-log-likelihood function is given by

$$J_{A_\mu | Y_v}(A) = -\text{trace}(B_{\mu|v}^T A) + \text{constant} \tag{111}$$

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. The MAP and MMSE attitude estimates for this PDF can be shown to be identical, but the conditional expectation of the attitude matrix is not an acceptable attitude estimate, since it is not an orthogonal matrix in general.[40] Equation (111) 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).[114] An attitude PDF of this form was first considered by Lo,[111-113] who referred to it as an exponential Fourier density. Lo included the higher-order $\ell = 2$ term, but treated neither process noise nor non-attitude bias parameters.

A PDF describing uncorrelated bias parameters obeying Gaussian statistics and an attitude with a PDF specified by Eqs. (110) and (111) would be the exponential of the sum of the right-hand-side of Eq. (111) 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

$$J_{A_\mu,\, \mathbf{x}_\mu | Y_v}(A,\, \mathbf{x}) = \frac{1}{2}\mathbf{x}^T F_{\mu|v}^x \mathbf{x} - \psi_{\mu|v}^T \mathbf{x} - \text{trace}[B_{\mu|v}^T(\mathbf{x})A] + \text{constant} \tag{112}$$

where

$$B_{\mu|v}(\mathbf{x}) \equiv B_{\mu|v,\,0} + \sum_{k=1}^{N} B_{\mu|v,\,k}\, x_k \tag{113}$$

where x_k is the kth component of $\mathbf{x}$. Propagation and update equations for the parameters $F_{t|\mu}^x$, $\psi_{t|\mu}$, and $B_{\mu|v,\,k}$ for $k = 0, \ldots, N$ are derived from the Fokker-Planck equation and Bayes' formula, respectively. It is not easy to establish that the MAP and MMSE estimates are equal for the correlated problem, and it may not even be true. Therefore, the method uses the more easily computed MAP estimates, which are found by simultaneously satisfying the equations

$$\hat{A}_{\mu/v} = \arg\min_{A \in \text{SO}(3)} \left[J_{A_\mu,\, \mathbf{x}_\mu | Y_v}(A,\, \hat{\mathbf{x}}_{\mu|v}) \right] \tag{114}$$

American Institute of Aeronautics and Astronautics

and

$$\hat{\mathbf{x}}_{\mu/v} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg \min} \left[J_{A_\mu, \mathbf{x}_\mu | Y_v}(\hat{A}_{\mu|v}, \mathbf{x}) \right] \tag{115}$$

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

$$[\boldsymbol{\eta}_{\mu|v}(A)]_k = \text{trace}(B_{\mu|v,\,k}^T A) \quad \text{for} \quad k = 1, \dots, N \tag{116}$$

allows us to rewrite Eq. (115) as

$$\begin{aligned}
\hat{\mathbf{x}}_{\mu/v} &= \underset{\mathbf{x} \in \mathbb{R}^N}{\arg \min} \left\{ \frac{1}{2}\mathbf{x}^T F_{\mu|v}^x \mathbf{x} - \left[\boldsymbol{\psi}_{\mu|v} + \boldsymbol{\eta}_{\mu|v}(\hat{A}_{\mu|v}) \right]^T \mathbf{x} - \text{trace}(B_{\mu|v,\,0}^T \hat{A}_{\mu|v}) \right\} \\
&= \left(F_{\mu|v}^x \right)^{-1} \left[\boldsymbol{\psi}_{\mu|v} + \boldsymbol{\eta}_{\mu|v}(\hat{A}_{\mu|v}) \right]
\end{aligned} \tag{117}$$

The attitude estimate $\hat{A}_{\mu|v}$ can be found by maximizing $\text{trace}[B_{\mu|v}^T(\hat{\mathbf{x}})A]$ for some initial guess for $\hat{\mathbf{x}}_{\mu|v}$, using one of the algorithms developed for Wahba's problem, then updating $\hat{\mathbf{x}}_{\mu|v}$ using Eq. (117), and iterating this procedure until it converges.

The Bayes measurement update equation and Fokker-Planck propagation equation for this filter were derived for an Earth-pointing spacecraft equipped with gyros and a three-axis magnetometer.[40] The filter developed an instability that is believed to result from the approximations in deriving the Fokker-Planck equation. Since this is a nonlinear equation, the $\ell = 1$ term in the PDF leads to $\ell = 2$ terms in the propagation, which cannot be accommodated in the filter. It is possible that including an $\ell = 2$ term in the PDF would stabilize the orthogonal filter, at the expense of increased complexity. 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 convergent 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[115]

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{f}[\hat{\mathbf{x}}(t)] + G[\hat{\mathbf{x}}(t)]\mathbf{d}(t) \tag{118a}$$

$$\hat{\mathbf{y}}(t) = \mathbf{h}[\hat{\mathbf{x}}(t)] \tag{118b}$$

where $\mathbf{d}(t)$ is the model error, which may include both model variations and external disturbances. A Taylor series expansion of the output estimate in Eq. (118b) is given by

$$\hat{\mathbf{y}}(t + \Delta t) = \hat{\mathbf{y}}(t) + \mathbf{z}[\hat{\mathbf{x}}(t), \Delta t] + \Lambda(\Delta t)S[\hat{\mathbf{x}}(t)]\mathbf{d}(t) \tag{119}$$

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

$$z_i[\hat{\mathbf{x}}(t), \Delta t] = \sum_{j=1}^{p_i} \frac{\Delta t^j}{j!} L_f^j(h_i) \tag{120}$$

where p_i, $i = 1, 2, \dots, m$, is the lowest order of the derivative of $h_i[\hat{\mathbf{x}}(t)]$ in which any component of $\mathbf{d}(t)$ first appears due to successive differentiation and substitution for $\dot{\hat{x}}_i(t)$ on the right side, and $L_f^j(h_i)$ is an ith order Lie derivative.[116] The matrix $\Lambda(\Delta t)$ is diagonal with elements given by $\Delta t^{p_i}/p_i!$, and the ith row of $S[\hat{\mathbf{x}}(t)]$ is given by

$$S_i[\hat{\mathbf{x}}(t)] = \left\{ L_{g_1}\left[L_f^{p_i-1}(h_i) \right], \dots, L_{g_q}\left[L_f^{p_i-1}(h_i) \right] \right\} \tag{121}$$

where q is the number of columns of $G[\hat{\mathbf{x}}(t)]$ and $L_{g_j}\left[L_f^{p_i-1}(h_i) \right]$ is another Lie derivative. Equation (121) 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 $\mathbf{d}(t)$. This yields[115, 117]

$$\mathbf{d}(t) = \left(\{ \Lambda(\Delta t) S[\hat{\mathbf{x}}(t)] \}^T R^{-1} \{ \Lambda(\Delta t) S[\hat{\mathbf{x}}(t)] \} + W \right)^{-1}$$
$$\times \{ \Lambda(\Delta t) S[\hat{\mathbf{x}}(t)] \}^T R^{-1} \{ \tilde{\mathbf{y}}(t + \Delta t) - \hat{\mathbf{y}}(t) - \mathbf{z}[\hat{\mathbf{x}}(t), \Delta t] \} \tag{122}$$

where the matrix W serves to weight the amount of model error added to correct the assumed model in Eq. (118). 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.[115] Equation (122) is used in Eq. (118a) 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 $\mathbf{d}(t)$ in $[t_k, t_{k+1}]$, and then the state estimates are propagated to time t_{k+1}.

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, since $\mathbf{d}(t)$ is used to propagate the quaternion kinematics directly. The predictive filter has been successfully tested using actual magnetometer and sun sensor data without gyros from the Solar, Anomalous, Magnetospheric Particle Explorer (SAMPEX) spacecraft.[118] Another advantage of the predictive filter is that it can be used to provide a deterministic solution by setting $W = 0$. For this approach the matrix inverse in Eq. (122) exists only if the minimum number of observations is present, e.g. two star measurements for three-axis attitude determination. This approach has been used to provide algorithms for the GPS attitude determination problem[119] and for the simultaneous attitude/position determination problem from multiple line-of-sight observations.[120] These papers show that the predictive filter solution for each of these problems achieves the associated Cramér-Rao lower bound, which proves that the predictive filter algorithms are *efficient* estimators.

X. Nonlinear Observers

Many nonlinear observers exist that provide attitude and angular rate estimates.[42–44, 46–48] Each has its various advantages and disadvantages. In this section we focus our attention to the observer designed by Thienel and Sanner,[48] since it is the most recent of the aforementioned references. In their approach, the measured angular rate of Eq. (14a) is rewritten as

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \boldsymbol{\beta} \tag{123}$$

where the vector $\boldsymbol{\beta}$ is now assumed to be constant. The estimate angular rate is given by $\hat{\boldsymbol{\omega}} = \tilde{\boldsymbol{\omega}} - \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the estimated bias. Observations of the true quaternion, $\mathbf{q}$, with no noise are assumed. The error quaternion between the "measured" quaternion and the estimated quaternion follows Eq. (20) with

$$\delta \mathbf{q} \equiv \begin{bmatrix} \delta \boldsymbol{\varrho} \\ \delta q_4 \end{bmatrix} = \mathbf{q} \otimes \hat{\mathbf{q}}^{-1} \tag{124}$$

The nonlinear observer for the quaternion and bias is given by

$$\dot{\hat{\mathbf{q}}} = \frac{1}{2} \Xi(\hat{\mathbf{q}}) A^T(\delta \mathbf{q}) [\hat{\boldsymbol{\omega}} + k \, \delta \boldsymbol{\varrho} \, \text{sign}(\delta q_4)] \tag{125a}$$

$$\dot{\hat{\boldsymbol{\beta}}} = -\frac{1}{2} \delta \boldsymbol{\varrho} \, \text{sign}(\delta q_4) \tag{125b}$$

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

$$\delta \dot{\mathbf{q}} = -\frac{1}{2} \Xi(\delta \mathbf{q}) [\Delta \boldsymbol{\beta} + k \, \delta \boldsymbol{\varrho} \, \text{sign}(\delta q_4)] \tag{126a}$$

$$\Delta \dot{\boldsymbol{\beta}} = \frac{1}{2} \delta \boldsymbol{\varrho} \, \text{sign}(\delta q_4) \tag{126b}$$

where $\Delta\beta \equiv \beta - \hat{\beta}$. The equilibrium states for Eq. (126) are $\delta\mathbf{q} = [0\ 0\ 0\ \pm 1]^T$ and $\Delta\beta = [0\ 0\ 0]^T$.

In order to prove global stability of the observer, the following candidate Lyapunov function is chosen:

$$V = \frac{1}{2}\Delta\beta^T\Delta\beta + \frac{1}{2}\begin{cases} (\delta q_4 - 1)^2 + \delta\varrho^T\delta\varrho, & \delta q_4 \geq 0 \\ (\delta q_4 + 1)^2 + \delta\varrho^T\delta\varrho, & \delta q_4 < 0 \end{cases} \tag{127}$$

Taking the time derivative of Eq. (127) and using Eq. (126) leads to

$$\dot{V} = -\frac{k}{2}\delta\varrho^T\delta\varrho \tag{128}$$

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

$$\ddot{V} = \frac{k}{2}\delta\varrho^T(\delta q_4 I_{3\times3} + [\delta\varrho\times])[\Delta\beta + k\,\delta\varrho\,\mathrm{sign}(\delta q_4)] \tag{129}$$

which is also bounded. Barbalat's lemma[121] then shows that $||\delta\varrho|| \to 0$ as $t \to \infty$.

Simulation results in Ref. 48 have shown that the nonlinear observer works well. Thienel and Sanner also have considered the effects of using the more realistic gyro model given by Eq. (14). The bias estimates can be shown to be exponentially convergent to a root-mean-square (RMS) bound given by

$$||\Delta\beta||_{\mathrm{RMS}} \leq \sqrt{3c}\sqrt{\frac{1}{4}\sigma_u^2 + \sigma_v^2} \tag{130}$$

where c is a finite constant which depends on the spacecraft inertia and desired rates. Coupling the observer with a nonlinear controller gives attitude tracking errors that converge asymptotically to the RMS bound

$$||\delta\varrho||_{\mathrm{RMS}} \leq \frac{\sqrt{3\gamma}}{k_D\lambda}\left[c^2\left(\frac{1}{4}\sigma_u^2 + \sigma_v^2\right) + \sigma_v^2\right]^{1/2} \tag{131}$$

where k_D and λ are control law gains, and γ is a bounded variable. Simulation results with noise show that the errors are within a ball determined by the variance of the noise.

Nonlinear observers are especially useful since they are often accompanied with global stability proofs. For spacecraft attitude estimation applications having the property of guaranteed convergence from any initial condition is especially desired by designers. Still, nonlinear observers are still in their infancy. The observers in Refs. 42–48 all require an attitude measurement, which limits their use to cases where a deterministic attitude is known. Still, these methods show great promise for future applications.

XI.   Adaptive Methods

Many adaptive methods exist that either update noise covariances in a filter design,[17,50–52] or update model parameters through least-squares techniques[54–57] or by using nonlinear techniques.[58–60] In this section we focus on a noise adaptive approach using a quaternion Kalman filter,[17] and a nonlinear adaptive approach to determine both inertia parameters and constant disturbance.[58]

A.   A Noise Adaptive Approach

The adaptive filter described in this section is based on a linear pseudo-measurement, 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 Ref. 17. The process noise covariance matrix is assumed to be a scalar times identity matrix $Q_k = \eta\,I_{3\times3}$, where η is the to-be-estimated parameter using an adaptive scheme. The process noise covariance is initialized using $Q_k = I_{3\times3}$. Denoting the sensitivity matrix that multiplies $\mathbf{q}$ in Eq. (25) as H, the i-step residual is given by

$$\nu_k^i = -H_k^i\,\hat{\mathbf{q}}_k^i \tag{132}$$

for $i = 1, 2, \ldots, m$, where H_k^i is computed from a single measurement and $\hat{\mathbf{q}}_k^i$ is the propagated quaternion estimate. Note that Eq. (132) implies that the pseudo-measurement is zero.[17] The sample mean of m predicted residuals is defined by

$$\bar{\nu}^m \equiv \frac{1}{m}\sum_{i=1}^{m}\nu_k^i \tag{133}$$

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

$$M^m = \frac{1}{m} \sum_{i=1}^{m} (\boldsymbol{\nu}_k^i)(\boldsymbol{\nu}_k^i)^T \tag{134}$$

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

$$\min_{\eta \geq 0} \left\{ J(\eta) = \| M^m - S^m(\eta) \|^2 \right\} \tag{135}$$

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

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

2. Propagate the quaternion and covariance using

$$\hat{\mathbf{q}}_k^i = \Phi_k^{i-1} \hat{\mathbf{q}}_k^{i-1} \tag{136a}$$

$$\mathcal{P}_k^i = (\Phi_k^{i-1}) \mathcal{P}_k^{i-1} (\Phi_k^{i-1})^T \tag{136b}$$

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

3. Compute $\Xi(\hat{\mathbf{q}}_k^i)$ and $\boldsymbol{\nu}_k^i$ from Eq. (132), along with the following variables

$$M^i = \frac{i-1}{i} M^{i-1} + \frac{1}{i} (\boldsymbol{\nu}_k^i)(\boldsymbol{\nu}_k^i)^T \tag{137a}$$

$$\mathcal{R}_k^i = \frac{1}{4} [\Xi(\hat{\mathbf{q}}_k^i)] R_k^i [\Xi(\hat{\mathbf{q}}_k^i)]^T + \alpha I_{4 \times 4} \tag{137b}$$

$$\mathcal{Z}^i = \left\{ (\Phi_k^{i-1}) [\Xi(\hat{\mathbf{q}}_k^{i-1})] [\Xi(\hat{\mathbf{q}}_k^{i-1})]^T (\Phi_k^{i-1})^T + [\Xi(\hat{\mathbf{q}}_k^i)] [\Xi(\hat{\mathbf{q}}_k^i)]^T \right\} \frac{\Delta t^2}{4} \tag{137c}$$

$$\mathcal{M}^i = \mathcal{M}^{i-1} + (H_k^i) \mathcal{P}_k^i (H_k^i)^T + \mathcal{R}_k^i \tag{137d}$$

$$L^i = L^{i-1} + (H_k^i) \mathcal{Z}_i (H_k^i)^T \tag{137e}$$

where R_k^i is the measurement covariance for the ith measurement and α is a small number in order 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

$$\hat{\eta} = \frac{\text{trace} \left[(M^m - \mathcal{M}^m)(L^m)^T \right]}{\text{trace} \left[(L^m)(L^m)^T \right]} \tag{138}$$

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. Simulation results from Ref. 17 show that the adaptive procedure is able to successfully update the covariance matrix for very large gyro bias errors.

B. A Nonlinear Adaptive Approach

The nonlinear adaptive approach described in this section estimates both the inertia matrix components and a constant vector of unknown disturbances.[58] The kinematics are given by the vector of MRPs,[62] denoted by $\mathbf{p}$, and Euler's dynamics equations:[122]

$$\dot{\mathbf{p}} = \frac{1}{4} \left\{ (1 - \mathbf{p}^T \mathbf{p}) I_{3 \times 3} + 2[\mathbf{p} \times] + 2\mathbf{p} \, \mathbf{p}^T \right\} \boldsymbol{\omega} \equiv \frac{1}{4} B(\mathbf{p}) \boldsymbol{\omega} \tag{139a}$$

$$\mathcal{J} \dot{\boldsymbol{\omega}} + [\boldsymbol{\omega} \times] \mathcal{J} \boldsymbol{\omega} = \mathbf{u} + \mathbf{F} \tag{139b}$$

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

$$\mathbf{u} = \begin{bmatrix} \hat{L} & \vdots & \hat{\mathcal{J}} & \vdots & \hat{\mathbf{F}} \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ \varphi \\ -1 \end{bmatrix} \equiv \hat{Q}\mathbf{x} \tag{140a}$$

$$\mathbf{g} \equiv [\omega_1^2 \ \ \omega_2^2 \ \ \omega_3^2 \ \ \omega_1\omega_2 \ \ \omega_2\omega_3 \ \ \omega_1\omega_3]^T \tag{140b}$$

$$\varphi = -K_v\omega - \left[\omega\,\omega^T + \left(\frac{4K_p}{1+||\mathbf{p}||^2} - \frac{||\omega||^2}{2} \right) I_{3\times3} \right] \mathbf{p} - 4K_i B^{-1}(\mathbf{p}) \int_0^t \mathbf{p} \ dt \tag{140c}$$

where $\hat{\mathcal{J}}$ is the estimated inertia matrix, $\hat{\mathbf{F}}$ 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

$$\hat{L}_1 \equiv \begin{bmatrix} 0 & \hat{\mathcal{J}}_{23} & -\hat{\mathcal{J}}_{23} \\ -\hat{\mathcal{J}}_{13} & 0 & \hat{\mathcal{J}}_{13} \\ \hat{\mathcal{J}}_{12} & -\hat{\mathcal{J}}_{12} & 0 \end{bmatrix}, \quad \hat{L}_2 \equiv \begin{bmatrix} \hat{\mathcal{J}}_{13} & \hat{\mathcal{J}}_{33} - \hat{\mathcal{J}}_{22} & -\hat{\mathcal{J}}_{12} \\ -\hat{\mathcal{J}}_{23} & \hat{\mathcal{J}}_{12} & \hat{\mathcal{J}}_{11} - \hat{\mathcal{J}}_{33} \\ \hat{\mathcal{J}}_{22} - \hat{\mathcal{J}}_{11} & -\hat{\mathcal{J}}_{13} & \hat{\mathcal{J}}_{23} \end{bmatrix} \tag{141}$$

The adaptive update law is given by

$$\dot{\hat{Q}} = -\frac{1}{4}B^T(\mathbf{p})S_3\mathbf{e}\,\mathbf{x}^T\Gamma \tag{142}$$

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

$$\mathbf{e} = \begin{bmatrix} \int_o^t (\mathbf{p} - \mathbf{p}_r) \ dt \\ \mathbf{p} - \mathbf{p}_r \\ \dot{\mathbf{p}} - \dot{\mathbf{p}}_r \end{bmatrix} \tag{143}$$

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

$$S\,E + E^T S = -D \tag{144}$$

where

$$E = \begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & I_{3\times3} \\ -K_i I_{3\times3} & -K_p I_{3\times3} & -K_v I_{3\times3} \end{bmatrix} \tag{145}$$

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. (141) 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. Due to the redundancy of the inertia terms, there is an infinity of solutions of these inertia terms which will all yield the desired behavior. Simulation results from Ref. 58 indicate that the adaptive control law is able to provide robust tracking in the presence of large system uncertainty.

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.

Sigma-point 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 Law of Large Numbers leads us to expect in all but pathological cases, and to be at worst unimodal. Sigma-point 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 sigma point 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 sigma-point 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 in the future, 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 pseudo-linear filters but are suboptimal compared to 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 other nonlinear filters discussed in this paper are mostly academic exercises, but they may provide the basis for applications as of now unforeseen. 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 deterministic estimation, but its advantages over the other filtering approaches have not been shown yet. Nonlinear observers are attractive since they usually are proven to be asymptotically stable, but due to their infancy they have not yet found widespread use on actual spacecraft yet. 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.

Acknowledgements

References

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

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

[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 Guidance and Control Conference*, Palo Alto, CA, Aug. 1978, AIAA-78-1248.

[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*, (in press).

[12]Bar-Itzhack, I. Y. and Oshman, Y., "Attitude Determination from Vector Observations: Quaternion Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-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 Guidance, Navigation, and Control Conference*, New Orleans, LA, Aug. 1991, AIAA-91-2706.

[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/CP-1992-3186) NASA-Goddard Space Flight Center, Greenbelt, 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., "A Novel Quaternion Filter," *AIAA Guidance, Navigation, and Control Conference*, Monterey, CA, Aug. 2002, AIAA-02-4460.

[18]Psiaki, M. L., "The Super-Iterated Extended Kalman Filter," *AIAA Guidance, Navigation, and Control Conference*, Providence, RI, Aug. 2004, AIAA-04-5418.

[19]Markley, F. L., Berman, N., and Shaked, U., "H_∞-Type Filter for Spacecraft Attitude Estimation," *AAS/GSFC International Symposium on Space Flight Dynamics*, Greenbelt, MD, May 1993, AAS-93-298.

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

[21]Smith, R. H., "An H_∞-Type Filter for GPS-Based Attitude Estimation," *AAS/AIAA Spaceflight Mechanics Conference*, Albuquerque, NM, Feb. 1995, AAS-95-134.

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

[23]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.

[24]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.

[25]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.

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

[27]Algrain, M. C. and Saniie, J., "Interlaced Kalman Filtering of 3-D Angular Motion Based on Euler's Nonlinear Equations," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-30, No. 1, 1994, pp. 174–184.

[28]Algrain, M. C. and Ehlers, D., "Novel Kalman Filtering Method for Suppression of Gyroscope Noise Effects in Pointing and Tracking Systems," *Optical Engineering*, Vol. 34, No. 10, 1995, pp. 3016–3030.

[29]Azor, R., Bar-Itzhack, I. Y., and Harman, R. R., "Satellite Angular Rate Estimation from Vector Measurements," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 3, 1998, pp. 450–457.

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

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

[32]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.

[33]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.

[34]Doucet, A., de Freitas, N., and Gordan, N., editors, *Sequential Monte Carlo Methods in Practice*, chap. 1, Springer, New York, NY, 2001.

[35]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.

[36]Cheng, Y. and Crassidis, J. L., "Particle Filtering for Sequential Spacecraft Attitude Estimation," *AIAA Guidance, Navigation, and Control Conference*, Providence, RI, Aug. 2004, AIAA-04-5337.

[37]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, Aug. 2004, AIAA-04-5340.

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

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

[40]Markley, F. L., "Attitude Filtering on SO(3)," *AAS Malcolm D. Shuster Astronautics Symposium*, Grand Island, NY, June 2005, AAS-05-460.

[41]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.

[42] 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*, London, England, May 1996, pp. 639–671.

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

[44] Salcudean, S., "A Globally Convergent Angular Velocity Observer for Rigid Body Motion," *IEEE Transactions on Automatic Control*, Vol. AC-36, No. 12, 1991, pp. 1493–1497.

[45] McDuffie, J. H. and Shtessel, Y. B., "A Sliding Mode Controller and Observer for Satellite Attitude Control," *AIAA Guidance, Navigation, and Control Conference*, New Orleans, LA, Aug. 1997, AIAA-97-3755.

[46] 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, Dayton, OH, July 1997, pp. 638–645.

[47] 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, Big Sky, MT, March 2000, pp. 505–511.

[48] 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. AC-48, No. 11, 2003, pp. 2011–2015.

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

[50] Ma, Z. and Ng, A., "Spacecraft Attitude Determination by Adaptive Kalman Filtering," *AIAA Guidance, Navigation, and Control Conference*, Monterey, CA, Aug. 2002, AIAA-02-4460.

[51] Lam, Q. M. and Wu, A., "Enhanced Precision Attitude Determination Algorithms," *Proceedings of the IEEE Aerospace Conference*, Vol. 1, Aspen, CO, March 1998, pp. 61–68.

[52] 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*, Albany, NY, Sept. 1995, pp. 176–181.

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

[54] Carter, M. T., Vadali, S. R., and Chamitoff, G. E., "Parameter Identification for the International Space Station Using Nonlinear Momentum Management Control," *AIAA Guidance, Navigation, and Control Conference*, New Orleans, LA, Aug. 1997, AIAA-97-3524.

[55] Kim, J.-W., Crassidis, J. L., Vadali, S. R., and Dershowitz, A. L., "International Space Station Leak Localization Using Vent Torque Estimation," *55th International Astronautical Conference*, Vancouver, BC, Oct. 2004, IAC-04-A.4.10.

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

[57] Kim, I., Kim, J., and Kim, Y., "Angular Rate Estimator using Disturbance Accommodation Technique," *AIAA Guidance, Navigation, and Control Conference*, Monterey, CA, Aug. 2002, AIAA-02-4829.

[58] 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.

[59] 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.

[60] Tanygin, S., "Generalization of Adaptive Attitude Tracking," *AIAA Guidance, Navigation, and Control Conference*, Monterey, CA, Aug. 2002, AIAA-02-4833.

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

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

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

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

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

[66] 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.

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

[68] Pandiyan, R., Solaiappan, A., and Malik, N., "A One Step Batch Filter for Estimating Gyroscope Calibration Parameters Using Star Vectors," *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, Providence, RI, Aug. 2004, AIAA-04-4858.

[69] Crassidis, J. L., "Sigma-Point Kalman Filtering for Integrated GPS and Inertial Navigation," *AIAA Guidance, Navigation, and Control Conference*, San Francisco, CA, Aug. 2005, AIAA-05-6052.

[70] 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.

[71] Sedlak, J. and Chu, D., "Kalman Filter Estimation of Attitude and Gyro Bias with the QUEST Observation Model," *AAS/GSFC International Symposium on Space Flight Dynamics*, Greenbelt, MD, May 1993, AAS-93-297.

[72] Markley, F. L., "Attitude Estimation or Quaternion Estimation?" *Journal of the Astronautical Sciences*, Vol. 52, No. 1/2, 2004, pp. 221–238.

[73]Cheng, Y., Crassidis, J. L., and Markley, F. L., "Attitude Estimation for Large Field-of-View Sensors," *AAS Malcolm D. Shuster Astronautics Symposium*, Grand Island, NY, June 2005, AAS-05-462.

[74]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," *6° Congresso Brasileiro de Automática*, Belo Horizonte, 1986, pp. 946–951.

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

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

[77]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.

[78]Toda, N. F., Heiss, J. L., and Schlee, F. H., "SPARS: the System, Algorithm, and Test Results," *Proceedings of the Symposium on Spacecraft Attitude Determination*, Vol. 1, Aerospace Corp. Report TR-0066 (5306)-12, Sept.-Oct. 1969, pp. 361–370.

[79]Yong, K. and Headley, R. P., "Real Time Precision Attitude Determination System (RETPAD) for Highly Maneuverable Spacecrafts," *AIAA Guidance and Control Conference*, Palo Alto, CA, Aug. 1978, AIAA-78-1246.

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

[81]Gray, C. W., "Star Tracker/IRU Attitude Determination Filters," *AAS Guidance and Control Conference*, Breckenridge, CO, Feb. 2001, AAS-01-039.

[82]Thompson, I. C. and Quasius, G. R., "Attitude Determination for the P80-1 Satellite," *AIAA Guidance and Control Conference*, Keystone, CO, Feb. 1980, AIAA-80-001.

[83]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.

[84]Paulson, D. C., Jackson, D. B., and Brown, C. D., "SPARS Algorithms and Simulation Results," *Proceedings of the Symposium on Spacecraft Attitude Determination*, Vol. 1, Aerospace Corp. Report TR-0066 (5306)-12, Sept.-Oct. 1969, pp. 293–317.

[85]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.

[86]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.

[87]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.

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

[89]Gelb, A., editor, *Applied Optimal Estimation*, The MIT Press, Cambridge, MA, 1974, pp. 190–191.

[90]Gill, P. E., Murray, W., and Wright, M. H., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, NY, 1981, pp. 88–93, 99–105, 133–141.

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

[92]Markley, F. L., "Attitude Dynamics," *Spacecraft Attitude Determination and Control*, edited by J. R. Wertz, chap. 16, Kluwer Academic Publishers, The Netherlands, 1978.

[93]Azor, R., Bar-Itzhack, I. Y., Deutschmann, J. K., and Harman, R. R., "Angular-Rate Estimation Using Delayed Quaternion Measurements," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 3, 2001, pp. 436–443.

[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*, The Johns Hopkins University Press, Baltimore, MD, 3rd ed., 1996, pp. 140–149, 601.

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

[97]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.

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

[99]Kim, S.-G., Crassidis, J. L., Cheng, Y., Fosbury, A. M., and Junkins, J. L., "Kalman Filtering for Relative Spacecraft Attitude and Position Estimation," *AIAA Guidance, Navigation, and Control Conference*, San Francisco, CA, Aug. 2005, AIAA-05-6087.

[100]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.

[101]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.

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

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

[104]Daum, F. and Huang, J., "Curse of Dimensionality and Particle Filters," *Proceedings of the IEEE Aerospace Conference*, Vol. 4, Big Sky, MT, March 2003, pp. 1979–1993.

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

[106]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.

[107]Oshman, Y. and Carmi, A., "Spacecraft Attitude Estimation from Vector Observations Using a Fast Particle Filter," *AAS/AIAA Space Flight Mechanics Conference*, Maui, HI, Feb. 2004, AAS-04-141.

[108]Oshman, Y. and Carmi, A., "Adaptive Estimation of Spacecraft Attitude from Vector Observations Using a Quaternion Particle Filter," *AAS Malcolm D. Shuster Astronautics Symposium*, Grand Island, NY, June 2005, AAS-05-464.

[109]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.

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

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

[112]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.

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

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

[115]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.

[116]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.

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

[118]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.

[119]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.

[120]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.

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

[122]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.