American Astronautical Society

ATTITUDE FILTERING ON SO(3)

F. Landis Markley

NASA Goddard Space Flight Center
Guidance, Navigation and Control Systems Eng. Branch
Greenbelt, MD 20771

The Malcolm D. Shuster Astronautics Symposium

ATTITUDE FILTERING ON SO(3)

F. Landis Markley*

A new method is presented for the simultaneous estimation of the attitude of a spacecraft and an N-vector of bias parameters. This method uses a probability distribution function defined on the Cartesian product of SO(3), the group of rotation matrices, and the Euclidean space $\mathbb{R}^N$. The Fokker-Planck equation propagates the probability distribution function between measurements, and Bayes's formula incorporates measurement update information. This approach avoids all the issues of singular attitude representations or singular covariance matrices encountered in extended Kalman filters. In addition, the filter has a consistent initialization for a completely unknown initial attitude, owing to the fact that SO(3) is a compact space.

INTRODUCTION

The extended Kalman filter (EKF) [1–7] 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 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 nonlinear filters, most recently sigma point or unscented filters [12–15] and particle filters [16–19]. These filters have been shown to perform better than the EKF in some cases at the expense of increased computational cost.

The optimal solution of the nonlinear estimation problem requires the propagation of the probability density function (PDF) of the state conditioned on the observation history [2, 19]. All practical nonlinear filters are approximations to this ideal. Exact finite dimensional filters [21] can be found that solve some nonlinear problems by using the Fokker-Planck (or forward Kolmogorov) equation [2, 20] to propagate a non-Gaussian PDF between measurements and Bayes's formula [1, 2] to incorporate measurement information. The filter we propose 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, 7, 9, 10] and unscented filters [15], 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. This approach is closely related to earlier work by Lo but has the advantage of including both process noise and parameters other than the attitude itself [22, 23].

*Guidance, Navigation and Control Systems Engineering Branch, NASA's Goddard Space Flight Center, Greenbelt, MD 20771,
landis.markley@nasa.gov

The paper begins with a review of some well-known facts about Gaussian PDFs to establish notation and to provide a background for the orthogonal filter. Then we introduce our parameterization for the attitude-only PDF and discuss some of its properties. With this preparation complete, we turn to the specification of the PDF for the combined estimation of attitude and an N-vector of bias parameters. We discuss the measurement update of the PDF by Bayes's formula and the propagation of the PDF between measurements by the Fokker-Planck equation, deriving explicit formulas for a specific measurement model and a specific dynamic propagation model. We then investigate the Cramér-Rao lower bound on the covariance, which gives a useful estimate of filter performance [1]. We test the new algorithm in a simulation of an Earth-pointing spacecraft equipped with gyros and a three-axis magnetometer, and conclude by summarizing the pros and cons of the orthogonal filter.

GAUSSIAN PDF

Let $\mathbf{x}_\mu$ denote an N-dimensional state vector at time t_μ and Y_ν denote the set of measurements up to and including time t_ν. Note that we will use Greek subscripts to denote time or measurement indices, to avoid confusion with Latin subscripts labeling matrix and vector components, and that we will be rather cavalier in distinguishing between random processes and their realizations. The Gaussian PDF with mean $\hat{\mathbf{x}}_{\mu|\nu}$ and positive definite covariance $P_{\mu|\nu}$ for $\mathbf{x}_\mu$ conditioned on the measurements Y_ν is

$$p_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x}) = [(2\pi)^N \det P_{\mu|\nu}]^{-1/2} \exp[-\tfrac{1}{2}(\mathbf{x}-\hat{\mathbf{x}}_{\mu|\nu})^T P_{\mu|\nu}^{-1}(\mathbf{x}-\hat{\mathbf{x}}_{\mu|\nu})] = \exp[-J_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})]. \tag{1}$$

The negative-log-likelihood function $J_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})$ is given by

$$J_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x}) \equiv \tfrac{1}{2}(\mathbf{x}-\hat{\mathbf{x}}_{\mu|\nu})^T P_{\mu|\nu}^{-1}(\mathbf{x}-\hat{\mathbf{x}}_{\mu|\nu}) + \ldots = \tfrac{1}{2}\mathbf{x}^T F_{\mu|\nu}^x \mathbf{x} - \phi_{\mu|\nu}^T \mathbf{x} + \ldots, \tag{2}$$

where $F_{\mu|\nu}^x \equiv P_{\mu|\nu}^{-1}$ is the positive definite information matrix and $\phi_{\mu|\nu} \equiv P_{\mu|\nu}^{-1}\hat{\mathbf{x}}_{\mu|\nu}$. The ellipses denote omitted terms that are independent of $\mathbf{x}$. The dependence on the measurement history Y_ν is not exhibited explicitly in the arguments of $p_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})$ and $J_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})$, but is implicitly indicated by the subscripts $\mu|\nu$ on the various parameters in the distribution. This PDF has $\tfrac{1}{2}N(N+3)$ free parameters, $\tfrac{1}{2}N(N+1)$ in the information matrix and N in the vector $\phi_{\mu|\nu}$, which contains information about the mean.

There are three estimates of interest: the conditional expectation

$$\overline{\mathbf{x}}_{\mu|\nu} \equiv \int_{\mathbb{R}^N} \mathbf{x}\, p_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})\, d^N\mathbf{x}, \tag{3}$$

the maximum *a posteriori* probability (MAP) estimate

$$\mathbf{x}_{\mu|\nu}^{\mathrm{MAP}} \equiv \begin{array}{c}\arg\max\\\mathbf{x}\in\mathbb{R}^N\end{array}\left[p_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})\right] = \begin{array}{c}\arg\min\\\mathbf{x}\in\mathbb{R}^N\end{array}\left[J_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})\right], \tag{4}$$

and the minimum mean-square error (MMSE) estimate

$$\mathbf{x}_{\mu|\nu}^{\mathrm{MMSE}} \equiv \begin{array}{c}\arg\min\\\mathbf{x}'\in\mathbb{R}^N\end{array}\left[\int_{\mathbb{R}^N}|\mathbf{x}-\mathbf{x}'|^2 p_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})\,d^N\mathbf{x}\right] = \begin{array}{c}\arg\min\\\mathbf{x}'\in\mathbb{R}^N\end{array}\left[\int_{\mathbb{R}^N}|\mathbf{x}-\overline{\mathbf{x}}_{\mu|\nu}|^2 p_{\mathbf{x}_\mu|Y_\nu}(\mathbf{x})\,d^N\mathbf{x} + |\overline{\mathbf{x}}_{\mu|\nu}-\mathbf{x}'|^2\right]. \tag{5}$$

It is easy to see that all three estimates are identical and are equal to $\hat{\mathbf{x}}_{\mu|\nu}$ for a Gaussian PDF. It is an interesting exercise to show that the Fokker-Planck equation and Bayes's formula lead to the usual Kalman filter propagation and measurement update for linear dynamics and measurement models, respectively [2].

ATTITUDE PDF

We define the PDF of the attitude $p_{A_\mu | Y_\nu}(A)$ so that $p_{A_\mu | Y_\nu}(A) \, dA$ is the probability that the attitude matrix is in some volume dA in SO(3) around A. The volume element dA is defined by the Haar measure on SO(3), which is unique up to a multiplicative constant [24, 25]. In parallel with the Gaussian case, we write the PDF in terms of a negative-log-likelihood function as

$$p_{A_\mu | Y_\nu}(A) = \exp[-J_{A_\mu | Y_\nu}(A)] . \tag{6}$$

The specification of the negative-log-likelihood function appeals to the Peter-Weyl Theorem, which states that a continuous function on a compact Lie group can be uniformly approximated by a linear combination of the matrix elements of the unitary irreducible representations of the group [26–28]. For the rotation group SO(3), these matrices are the $(2\ell + 1) \times (2\ell + 1)$ rotation matrices D^ℓ for integer ℓ [27, 29], so we have

$$J_{A_\mu | Y_\nu}(A) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} c_{m'm}^\ell D_{m'm}^\ell , \tag{7}$$

where the $c_{m'm}^\ell$ are scalar coefficients. For numerical work, this series on SO(3) must be truncated at some finite ℓ_{max} in the same sense that equation (2) can be regarded as a Taylor series in $\mathbf{x}$ on $\mathbb{R}^N$ truncated at order 2. A PDF in the form of equation (7), truncated at $\ell_{max} = 2$, was first considered by Lo [22, 23], who referred to it as an exponential Fourier density. We will truncate the series even more drastically at $\ell_{max} = 1$, but we will include both process noise and non-attitude bias parameters, neither of which was considered by Lo. The appendix contains some preliminary thoughts on extending the representation to $\ell = 2$. We prefer not to deal with the complex matrices D^ℓ, so we write the truncated equation (7) as

$$J_{A_\mu | Y_\nu}(A) = -\operatorname{tr}(B_{\mu|\nu}^T A) + \ldots , \tag{8}$$

where $B_{\mu|\nu}$ is a real 3×3 matrix and the ellipses denote terms independent of A that are determined in principle by the normalization of the PDF. We will see that it is not necessary to compute these terms.

The unitary equivalence of equations (7) and (8) is established by the relations [29] $D^0 = 1$ and

$$D^1(A) = M_1^\dagger A M_1 , \tag{9}$$

where M_1 is the unitary matrix defined by

$$M_1 \equiv \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -i/\sqrt{2} & 0 & -i/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} , \tag{10}$$

with the dagger denoting the Hermitian conjugate (transpose complex conjugate) matrix.

All the information about the PDF is contained in the real 3×3 matrix $B_{\mu|\nu}$, which has been known for some time to contain the correct number of free parameters to represent the mean and covariance of a state with three degrees of freedom, namely three parameters for the mean and six for the attitude covariance [30]. In fact, a filter based on equation (8) will look very much like Shuster's Filter QUEST [31]. The representation of the PDF for spacecraft attitude given by equations (6) and (8) is completely free of singularities, discontinuities, or constraints. Shuster has pointed out another advantage of defining the PDF on the compact manifold of SO(3) rather than on some 3-dimensional Euclidean tangent space, namely that it allows us to represent a complete lack of attitude knowledge by a uniform probability distribution [32].

It is possible to define three estimates in parallel with the Gaussian case. The conditional expectation

$$\overline{A}_{\mu|\nu} \equiv \int_{SO(3)} A \, p_{A_\mu|Y_\nu}(A) \, dA \tag{11}$$

does not provide a valid attitude estimate, however, because it is not an orthogonal matrix in general, i.e. it does not belong to SO(3). This has led to the erroneous belief that PDFs on non-Euclidean manifolds necessarily yield unphysical estimates [11]. In contrast to this, Oshman and Carmi [19] have emphasized that perfectly valid attitude estimates are provided by the MAP estimate

$$A_{\mu|\nu}^{MAP} \equiv \underset{A \in SO(3)}{\arg\max} \left[p_{A_\mu|Y_\nu}(A) \right] = \underset{A \in SO(3)}{\arg\max} \left[\mathrm{tr}(B_{\mu|\nu}^T A) \right] \tag{12}$$

and the MMSE estimate

$$A_{\mu|\nu}^{MMSE} \equiv \underset{A' \in SO(3)}{\arg\min} \left[\int_{SO(3)} \|A - A'\|_F^2 \, p_{A_\mu|Y_\nu}(A) \, dA \right]$$

$$= \underset{A' \in SO(3)}{\arg\min} \left[\int_{SO(3)} \|A - \overline{A}_{\mu|\nu}\|_F^2 \, p_{A_\mu|Y_\nu}(A) \, dA + \|\overline{A}_{\mu|\nu} - A'\|_F^2 \right] = \underset{A' \in SO(3)}{\arg\min} \|\overline{A}_{\mu|\nu} - A'\|_F^2$$

$$= \underset{A' \in SO(3)}{\arg\min} \left[\|\overline{A}_{\mu|\nu}\|_F^2 - 2\,\mathrm{tr}(\overline{A}_{\mu|\nu}^T A') + \|I_{3\times3}\|_F^2 \right] = \underset{A' \in SO(3)}{\arg\max} \left[\mathrm{tr}(\overline{A}_{\mu|\nu}^T A') \right]. \tag{13}$$

Equation (13) employs the Frobenius norm, defined by

$$\|M\|_F^2 \equiv \mathrm{tr}(MM^T) = \sum_{i,j} M_{ij}^2 . \tag{14}$$

There are many ways to find the MAP and MMSE estimates [33]. We will use the SVD method [34, 35], which is not the most computationally efficient, but has excellent numerical stability and provides the greatest analytical insight. We find the MAP by performing the SVD of the matrix $B_{\mu|\nu}$

$$B_{\mu|\nu} = USV^T = U_+ S' V_+^T , \tag{15}$$

where S is a diagonal matrix with non-negative elements in descending order of magnitude down the main diagonal, and where U and V are orthogonal, but not necessarily proper, matrices. The matrices in the right-hand member of equation (15) are defined by

$$U_+ \equiv U \, \mathrm{diag}([1, 1, \det U]), \quad V_+ \equiv V \, \mathrm{diag}([1, 1, \det V]), \quad \text{and} \quad S' \equiv S \, \mathrm{diag}([1, 1, \det U \det V]), \tag{16}$$

so that U_+ and V_+ are proper orthogonal matrices. Then the SVD method tells us that

$$A_{\mu|\nu}^{MAP} = U_+ V_+^T = U \, \mathrm{diag}([1, 1, \det U \det V]) V^T . \tag{17}$$

The SVD method will return an attitude estimate for any $B_{\mu|\nu}$, but $B_{\mu|\nu}$ must have rank greater than one for the estimate to be valid. The MMSE estimate is found similarly from the SVD of $\overline{A}_{\mu|\nu}$:

$$\overline{A}_{\mu|\nu} = \overline{U}\overline{S}\overline{V}^T = \overline{U}_+ \overline{S}' \overline{V}_+^T \tag{18}$$

and

$$A_{\mu|\nu}^{MMSE} = \overline{U}_+ \overline{V}_+^T . \tag{19}$$

We will now show that the MAP and MMSE estimates are identical in this case. Equations (8) and (15) give

$$J_{A_\mu | Y_V}(A) = -\mathrm{tr}(V_+ S' U_+^T A) + \ldots = -\mathrm{tr}(S' U_+^T A V_+) + \ldots = -\mathrm{tr}(S' A') + \ldots \qquad (20)$$

with

$$A' \equiv U_+^T A V_+ \qquad (21)$$

Then equation (11) gives

$$\bar{A}_{\mu|v} = \int_{SO(3)} A \exp[-J_{A_\mu | Y_V}(A)]\,dA = U_+ \left\{ \int_{SO(3)} A' \exp[\mathrm{tr}(S'A') + \ldots]\,dA' \right\} V_+^T . \qquad (22)$$

The invariance of the Haar measure allowed us to change the integration over dA to integration over dA'. We can use any convenient parameterization for A' to perform the integral, and we choose to use a quaternion [36, 37] with vector part $\mathbf{q}$ and scalar part $\sqrt{1 - |\mathbf{q}|^2}$;

$$A'(\mathbf{q}) = \begin{bmatrix} 1 - 2(q_2^2 + q_3^2) & 2\left(q_1 q_2 + q_3 \sqrt{1-|\mathbf{q}|^2}\right) & 2\left(q_1 q_3 - q_2 \sqrt{1-|\mathbf{q}|^2}\right) \\ 2\left(q_1 q_2 - q_3 \sqrt{1-|\mathbf{q}|^2}\right) & 1 - 2(q_3^2 + q_1^2) & 2\left(q_2 q_3 + q_1 \sqrt{1-|\mathbf{q}|^2}\right) \\ 2\left(q_1 q_3 + q_2 \sqrt{1-|\mathbf{q}|^2}\right) & 2\left(q_2 q_3 - q_1 \sqrt{1-|\mathbf{q}|^2}\right) & 1 - 2(q_1^2 + q_2^2) \end{bmatrix} . \qquad (23)$$

The square root singularity in this parameterization is of no concern, since the quaternion is only used as a dummy variable of integration, not as the filter's attitude representation. The volume element in the quaternion parameterization, normalized so that SO(3) has unit volume, is [32]

$$dA' = \pi^{-2}(1 - |\mathbf{q}|^2)^{-1/2}\, d^3\mathbf{q} . \qquad (24)$$

It is clear from inspection of equation (23) that all the off-diagonal elements of the matrix integral in equation (22) vanish by symmetry in $\mathbf{q}$. Comparison of equations (18) and (22) then shows that

$$\bar{U}_+ = U_+, \quad \bar{V}_+ = V_+, \quad \text{and} \quad \bar{S}'_{ii} = \pi^{-2} \int_{|\mathbf{q}|^2 \leq 1} A'_{ii} \exp[\mathrm{tr}(S'A') + \ldots](1 - |\mathbf{q}|^2)^{-1/2}\, d^3\mathbf{q} \quad \text{for} \quad i = 1, 2, 3 \quad (25)$$

and thus that equations (17) and (19) give

$$A_{\mu|v}^{MMSE} = A_{\mu|v}^{MAP} . \qquad (26)$$

Note that the normalization of the PDF enters into $\bar{S}'$, but not into the attitude estimate.

ATTITUDE/PARAMETER PDF

We now turn to the problem of simultaneously estimating the attitude and an N-dimensional vector of bias parameters. If these were independent, their joint PDF would just be the product of the PDFs of equations (1) and (6), so the negative-log likelihood function would be the sum of equations (2) and (8). This PDF would have $\frac{1}{2}N(N+3) + 9$ free parameters. The joint PDF won't have this simple form when the attitude and bias parameters are correlated, of course. An EKF for this problem would have $\frac{1}{2}(N+3)(N+6)$ or $\frac{1}{2}N(N+3) + 3N + 9$ parameters to describe the mean and covariance of the state with $N+3$ independent components. The additional $3N$ parameters contain information about the correlations between the three attitude parameters and the N bias parameters.

The terms in a Gaussian EKF describing the correlations between the attitude parameters and the bias parameters are linear in the components of the bias parameter vector and in the components of the attitude parameter vector. The obvious generalization for the orthogonal filter is to include terms in the joint PDF that are linear in both the bias vector and the attitude matrix, which leads to the *ansatz*

$$J_{A_\mu, \mathbf{x}_\mu | Y_\nu}(A, \mathbf{x}) = \tfrac{1}{2} \mathbf{x}^T F_{\mu|\nu}^x \mathbf{x} - \phi_{\mu|\nu}^T \mathbf{x} - \text{tr}[B_{\mu|\nu}^T(\mathbf{x})A] + \dots, \tag{27}$$

where

$$B_{\mu|\nu}(\mathbf{x}) \equiv B_{\mu|\nu,0} + \sum_{k=1}^{N} B_{\mu|\nu,k} x_k . \tag{28}$$

As before, the ellipses denote terms that are independent of the state variables, i.e. of both A and $\mathbf{x}$. Each $B_{\mu|\nu,k}$ for $k = 0, \dots, N$ is a 3×3 matrix independent of the state variables, so we have introduced $6N$ more parameters than are used by an EKF. This can be viewed as either an advantage, in that the orthogonal filter carries more information about correlations, or a disadvantage, since it increases the computational effort. The number of extra parameters for $1 \le N \le 18$ obeys the relationship

$$\frac{3}{7} \le \frac{\text{number of extra parameters in the orthogonal filter}}{\text{number of parameters in an EKF}} \le \frac{24}{35} . \tag{29}$$

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, we will concentrate on the more easily computed MAP estimates, which we will denote by carets. These are found by simultaneously satisfying the equations

$$\hat{A}_{\mu|\nu} = \genfrac{}{}{0pt}{}{\arg\min}{A \in SO(3)} \left[J_{A_\mu, \mathbf{x}_\mu | Y_\nu}(A, \hat{\mathbf{x}}_{\mu|\nu}) \right] \tag{30}$$

and

$$\hat{\mathbf{x}}_{\mu|\nu} = \genfrac{}{}{0pt}{}{\arg\min}{\mathbf{x} \in \mathbb{R}^N} \left[J_{A_\mu, \mathbf{x}_\mu | Y_\nu}(\hat{A}_{\mu|\nu}, \mathbf{x}) \right], \tag{31}$$

with $J_{A_\mu, \mathbf{x}_\mu | Y_\nu}(A, \mathbf{x}_{\mu|\nu})$ given by equations (27) and (28). The attitude estimate is given by equations (15) and (17), remembering to use $B_{\mu|\nu}(\hat{\mathbf{x}}_{\mu|\nu})$. Defining a vector function $\boldsymbol{\eta}_{\mu|\nu}(A)$ with components given by

$$[\boldsymbol{\eta}_{\mu|\nu}(A)]_k = \text{tr}(B_{\mu|\nu,k}^T A) \quad \text{for} \quad k = 1, \dots, N . \tag{32}$$

allows us to rewrite equation (31) as

$$\hat{\mathbf{x}}_{\mu|\nu} = \genfrac{}{}{0pt}{}{\arg\min}{\mathbf{x} \in \mathbb{R}^N} \left\{ \tfrac{1}{2} \mathbf{x}^T F_{\mu|\nu}^x \mathbf{x} - [\phi_{\mu|\nu} + \boldsymbol{\eta}_{\mu|\nu}(\hat{A}_{\mu|\nu})]^T \mathbf{x} - \text{tr}(B_{\mu|\nu,0}^T \hat{A}_{\mu|\nu}) + \dots \right\}, \tag{33}$$

so the MAP estimate of the parameter vector is

$$\hat{\mathbf{x}}_{\mu|\nu} = (F_{\mu|\nu}^x)^{-1} [\phi_{\mu|\nu} + \boldsymbol{\eta}_{\mu|\nu}(\hat{A}_{\mu|\nu})] . \tag{34}$$

This problem can be solved by finding $\hat{A}_{\mu|\nu}$ from the SVD of $B_{\mu|\nu}(\hat{\mathbf{x}}_{\mu|\nu})$ using some initial guess for $\hat{\mathbf{x}}_{\mu|\nu}$, then updating $\hat{\mathbf{x}}_{\mu|\nu}$ using equations (32) and (34), and iterating this procedure until it converges. The literature may suggest better ways to find this simultaneous minimum, however [38–40].

6

MEASUREMENT UPDATE

We will consider the measurement update before we consider the time propagation between measurements, since it is somewhat simpler. We assume that we have propagated the state to the time t_μ using measurements up to the previous time $t_{\mu-1}$. The update of the PDF to incorporate the information in the measurement $\mathbf{y}_\mu$ is given by Bayes's theorem for probability densities [1, 2]

$$p_{A_\mu,\mathbf{x}_\mu|Y_\mu}(A,\mathbf{x}) = \frac{p_{\mathbf{y}_\mu|A_\mu,\mathbf{x}_\mu}(\mathbf{y}_\mu|A,\mathbf{x})\,p_{A_\mu,\mathbf{x}_\mu|Y_{\mu-1}}(A,\mathbf{x})}{\int\limits_{SO(3)}\int\limits_{\mathbb{R}^N} p_{\mathbf{y}_\mu|A_\mu,\mathbf{x}_\mu}(\mathbf{y}_\mu|A,\mathbf{x})\,p_{A_\mu,\mathbf{x}_\mu|Y_{\mu-1}}(A,\mathbf{x})\,d^N\mathbf{x}\,dA}, \tag{35}$$

where $p_{\mathbf{y}_\mu|A_\mu,\mathbf{x}_\mu}(\mathbf{y}_\mu|A,\mathbf{x})$ is the PDF for the measurement conditioned on the pre-measurement state. The denominator of equation (35) only contributes terms independent of A and $\mathbf{x}$, which are not needed to compute the MAP, so the update in terms of the negative-log-likelihood function is

$$J_{A_\mu,\mathbf{x}_\mu|Y_\mu}(A,\mathbf{x}) = J_{A_\mu,\mathbf{x}_\mu|Y_{\mu-1}}(A,\mathbf{x}) - \ln[p_{\mathbf{y}_\mu|A_\mu,\mathbf{x}_\mu}(\mathbf{y}_\mu|A,\mathbf{x})] + \ldots, \tag{36}$$

with the ellipses standing for omitted terms that are independent of A and $\mathbf{x}$.

The update can be performed exactly if the functional dependence on A and $\mathbf{x}$ of the right side of equation (36) is the same as the functional dependence of equation (27). If this is not the case, an approximation will be required. We will consider a specific example, but the method outlined can be extended in principle to quite general measurement models. Consider a measurement $\mathbf{y}_\mu$ of the ambient magnetic field by a triaxial magnetometer with a constant bias $\mathbf{x}_m$ in the spacecraft body frame and with Gaussian measurement errors with measurement information matrix (inverse covariance) F_μ^m. For this measurement model,

$$-\ln[p_{\mathbf{y}_\mu|A_\mu,\mathbf{x}_\mu}(\mathbf{y}_\mu|A,\mathbf{x})] = \tfrac{1}{2}(\mathbf{y}_\mu - A\mathbf{H}_\mu - \mathbf{x}_m)^T F_\mu^m (\mathbf{y}_\mu - A\mathbf{H}_\mu - \mathbf{x}_m)$$
$$= \tfrac{1}{2}\mathbf{y}_\mu^T F_\mu^m \mathbf{y}_\mu + \tfrac{1}{2}\mathbf{x}_m^T F_\mu^m \mathbf{x}_m - \mathbf{y}_\mu^T F_\mu^m \mathbf{x}_m - (\mathbf{y}_\mu - \mathbf{x}_m)^T F_\mu^m A\mathbf{H}_\mu + \tfrac{1}{2}\mathbf{H}_\mu^T A^T F_\mu^m A\mathbf{H}_\mu, \tag{37}$$

where $\mathbf{H}_\mu$ is the magnetic field in the coordinate frame to which the attitude is referenced. All the terms on the right side of this equation except the last are of the desired form. The last term is actually independent of A and $\mathbf{x}$ if F_μ^m is a multiple of the identity matrix. If F_μ^m is not a multiple of the identity matrix this term has $\ell = 0$, $\ell = 1$, and $\ell = 2$ parts [29]. Since the PDF does not contain an $\ell = 2$ part, we must discard this part of the last term on the right side of equation (37), giving the approximation

$$\tfrac{1}{2}\mathbf{H}_\mu^T A^T F_\mu^m A\mathbf{H}_\mu \approx \psi_\mu + \mathrm{tr}(\Psi_\mu^T A). \tag{38}$$

We evaluate the $\ell = 0$ part ψ_μ and the $\ell = 1$ part Ψ_μ by means of the relationships [32]

$$\int\limits_{SO(3)} dA = 1, \quad \int\limits_{SO(3)} A\,dA = 0, \quad \int\limits_{SO(3)} A_{ij}A_{k\ell}\,dA = \tfrac{1}{3}\delta_{ik}\delta_{j\ell}, \quad \text{and} \quad \int\limits_{SO(3)} A_{ij}A_{k\ell}A_{mn}\,dA = \tfrac{1}{6}\varepsilon_{ikm}\varepsilon_{j\ell n}. \tag{39}$$

The first of these equations follows from the normalization of the volume of SO(3), and the second reflects the orthogonality of the $\ell = 0$ and $\ell = 1$ irreducible representations of SO(3). The third uses the Kronecker delta symbol δ_{ik} and the fourth uses the Levi-Civita antisymmetric symbol ε_{ikm}, defined by $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$, $\varepsilon_{132} = \varepsilon_{321} = \varepsilon_{213} = -1$, $\varepsilon_{ijk} = 0$ if any two indices are equal. This gives

$$\psi_\mu = \int\limits_{SO(3)} [\psi_\mu + \mathrm{tr}(\Psi_\mu^T A)]\,dA = \tfrac{1}{2}\int\limits_{SO(3)} (\mathbf{H}_\mu^T A^T F_\mu^m A\mathbf{H}_\mu)\,dA = \tfrac{1}{6}|\mathbf{H}_\mu|^2\,\mathrm{tr}\,F_\mu^m \tag{40}$$

and

$$\Psi_\mu = 3\int\limits_{SO(3)} A[\psi_\mu + \mathrm{tr}(\Psi_\mu^T A)]\,dA = \tfrac{3}{2}\int\limits_{SO(3)} A(\mathbf{H}_\mu^T A^T F_\mu^m A\mathbf{H}_\mu)\,dA = 0. \tag{41}$$

The vanishing of Ψ_μ results from the antisymmetry of the Levi-Civita symbol and the symmetry of F_μ^m.

Substituting equation (37) into equation (36) and inserting equation (27) gives measurement update

$$\frac{1}{2}\mathbf{x}^T F_{\mu|\mu}^x \mathbf{x} - \phi_{\mu|\mu}^T \mathbf{x} - \text{tr}[B_{\mu|\mu}^T(\mathbf{x})A] = \frac{1}{2}\mathbf{x}^T F_{\mu|\mu-1}^x \mathbf{x} - \phi_{\mu|\mu-1}^T \mathbf{x} - \text{tr}[B_{\mu|\mu-1}^T(\mathbf{x})A] + \frac{1}{2}\mathbf{x}_m^T F_\mu^m \mathbf{x}_m$$
$$- \mathbf{y}_\mu^T F_\mu^m \mathbf{x}_m - (\mathbf{y}_\mu - \mathbf{x}_m)^T F_\mu^m A \mathbf{H}_\mu + \ldots, \tag{42}$$

where the ellipses indicating neglected terms independent of A (i.e., having $\ell = 0$) and $\mathbf{x}$. The explicit update of the parameters in the PDF, assuming that the components of $\mathbf{x}_m$ are the first three components in the bias vector $\mathbf{x}$, follows from this:

$$F_{\mu|\mu}^x = F_{\mu|\mu-1}^x + \begin{bmatrix} F_\mu^m & 0_{3\times(N-3)} \\ 0_{(N-3)\times 3} & 0_{(N-3)\times(N-3)} \end{bmatrix} \tag{43}$$

$$\phi_{\mu|\mu} = \phi_{\mu|\mu-1} + \begin{bmatrix} F_\mu^m \mathbf{y}_\mu \\ \mathbf{0}_{N-3} \end{bmatrix} \tag{44}$$

$$B_{\mu|\mu,0} = B_{\mu|\mu-1,0} + F_\mu^m \mathbf{y}_\mu \mathbf{H}_\mu^T . \tag{45}$$

$$B_{\mu|\mu,i} = B_{\mu|\mu-1,i} - F_\mu^m \mathbf{e}_i \mathbf{H}_\mu^T . \qquad\qquad \text{for } i = 1, 2, 3 \tag{46a}$$

$$B_{\mu|\mu,i} = B_{\mu|\mu-1,i} \qquad\qquad \text{for } i > 3, \tag{46b}$$

where $\mathbf{e}_i$ denotes the coordinate axis unit vectors:

$$\mathbf{e}_1 \equiv [1\,0\,0]^T, \quad \mathbf{e}_2 \equiv [0\,1\,0]^T, \quad \mathbf{e}_3 \equiv [0\,0\,1]^T . \tag{47}$$

In the attitude-only case, the only update is given by equation (45), which is the same as in Shuster's Filter QUEST algorithm [31].

PROPAGATION BETWEEN MEASUREMENTS

We now turn to the procedure for propagating the PDF from time t_μ to time $t_{\mu+1}$, i.e. for propagating $p_{A_\mu, \mathbf{x}_\mu | Y_\mu}(A, \mathbf{x})$ to $p_{A_{\mu+1}, \mathbf{x}_{\mu+1} | Y_\mu}(A, \mathbf{x})$. We use the Fokker-Planck partial differential equation (PDE) for the PDF, or actually an equivalent PDE for the negative-log-likelihood function, to derive ordinary differential equations for the parameters $F_{t|\mu}^x$, $\phi_{t|\mu}$, and $B_{t|\mu,i}$ in the PDF. The subscripts $t \mid \mu$ are to convey the idea that these quantities will be integrated over t from t_μ to $t_{\mu+1}$. We will, however, suppress many subscripts and arguments throughout this section to simplify the notation.

We begin by stacking the columns $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ of the attitude matrix A in a nine-component vector $\mathbf{x}_A$

$$A = [\mathbf{u}\,\mathbf{v}\,\mathbf{w}] \quad \text{and} \quad \mathbf{x}_A \equiv \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{bmatrix} . \tag{48}$$

Then we define the (N+9)-component state vector ξ as

$$\xi \equiv \begin{bmatrix} \mathbf{x}_A \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_A \\ \mathbf{x}_m \\ \mathbf{x}_d \end{bmatrix}, \tag{49}$$

where $\mathbf{x}_m$ is the vector of measurement biases and $\mathbf{x}_d$ is a vector of dynamic biases. The Fokker-Planck equation is derived from the stochastic differential equation (SDE) [2, 20, 41] obeyed by the state vector ξ

$$d\xi = \mathbf{f}\,dt + G\,d\mathbf{w}, \tag{50}$$

where $d\mathbf{w}$ is a Wiener process. Equation (50) should be considered as a Stratonovich SDE rather than as an Itô SDE for two reasons. First, solutions of a Stratonovich SDE, unlike those of an Itô SDE, obey the usual laws of calculus, including the product rule for differentiation. Thus the orthogonality constraint on A is preserved by a Stratonovich SDE, but would be violated by an Itô SDE. The second argument appeals to the Wong-Zakai theorem [2, 41], which says that, as a rule of thumb, the Stratonovich SDE represents the unphysical white-noise infinite-bandwidth limit of finite-bandwidth noise. It is reassuring that all these issues with SDEs can be forgotten after we have derived the Fokker-Planck equation.

The Fokker-Planck equation arising from the Stratonovich SDE given by equation (50) is [20]

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{N+9} \frac{\partial (f_i p)}{\partial \xi_i} + \frac{1}{2} \sum_{i,j,k=1}^{N+9} \frac{\partial}{\partial \xi_i} \left[G_{ik} \frac{\partial (G_{jk} p)}{\partial \xi_j} \right]. \tag{51}$$

Substituting $p = \exp(-J)$ gives the PDE for the negative-log-likelihood function

$$\frac{\partial J}{\partial t} = \sum_{i=1}^{N+9} \left[\frac{\partial f_i}{\partial \xi_i} - f_i \frac{\partial J}{\partial \xi_i} - \frac{1}{2} \left(\frac{\partial g_i}{\partial \xi_i} - g_i \frac{\partial J}{\partial \xi_i} \right) \right], \tag{52}$$

where the $(N+9)$-component vector $\mathbf{g}$ is defined by

$$g_i \equiv \sum_{j,k=1}^{N+9} G_{ik} \left(\frac{\partial G_{jk}}{\partial \xi_j} - G_{jk} \frac{\partial J}{\partial \xi_j} \right). \tag{53}$$

The left side of equation (52) is easy to evaluate; it is simply

$$\frac{\partial J}{\partial t} = \frac{1}{2} \mathbf{x}^T \dot{F}_{t|\mu}^x \mathbf{x} - \dot{\boldsymbol{\phi}}_{t|\mu}^T \mathbf{x} - \mathrm{tr}(\dot{B}_{t|\mu,0}^T A) - \sum_{k=1}^{N} x_k \mathrm{tr}(\dot{B}_{t|\mu,k}^T A) + \ldots, \tag{54}$$

where the dots denote time derivatives and the ellipses, as always, indicate terms independent of A and $\mathbf{x}$ that do not concern us. Evaluating the right side of equation (52) requires the specification of a dynamic model for $\boldsymbol{\xi}$. We will choose the model of a spacecraft provided with three-axis rate information from gyros satisfying Farrenkopf's noise model, with the gyro dynamics used in place of an Eulerian dynamics model of rigid body motion [6, 42]. The techniques we apply to this particular model can be generalized to other dynamic models, as was the case for measurement modeling.

The attitude matrix obeys the SDE

$$dA = -[d\boldsymbol{\theta} \times] A \tag{55}$$

where the cross-product matrix is defined as

$$[\mathbf{r} \times] \equiv \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix}. \tag{56}$$

and the infinitesimal rotation vector is given by

$$d\boldsymbol{\theta} = (\boldsymbol{\omega}_g - \mathbf{x}_d) dt - G_v d\mathbf{w}_v. \tag{57}$$

The three-dimensional vectors $\boldsymbol{\omega}_g$ and $\mathbf{x}_d$ are the gyro rate output vector and the dynamic bias vector, a vector of gyro drift biases for this dynamic model. The Wiener process $d\mathbf{w}_v$ is the gyro drift rate noise process. The gyro drift rate biases are assumed to obey the SDE

$$d\mathbf{x}_d = G_u d\mathbf{w}_u, \tag{58}$$

9

where $d\mathbf{w}_u$ is the gyro drift rate ramp noise Wiener process. Using equation (48), equations (55) and (57) can be put in the vector form

$$d\mathbf{x}_A = -\Gamma_A^T d\boldsymbol{\theta} = \Gamma_A^T[(\mathbf{x}_d - \boldsymbol{\omega}_g)dt + G_v d\mathbf{w}_v], \tag{59}$$

where

$$\Gamma_A \equiv \begin{bmatrix} [\mathbf{u}\times] & [\mathbf{v}\times] & [\mathbf{w}\times] \end{bmatrix}. \tag{60}$$

Now we assume that the measurement bias parameters obey an equation analogous to equation (58)

$$d\mathbf{x}_m = G_m d\mathbf{w}_m, \tag{61}$$

to arrive at a state dynamics equation in the form of equation (50) with

$$\mathbf{f} \equiv \begin{bmatrix} \Gamma_A^T(\mathbf{x}_d - \boldsymbol{\omega}_g) \\ \mathbf{0}_{N-3} \\ \mathbf{0}_3 \end{bmatrix}, \quad G = \begin{bmatrix} \Gamma_A^T G_v & 0_{3\times(N-3)} & 0_{3\times3} \\ 0_{(N-3)\times3} & G_m & 0_{(N-3)\times3} \\ 0_{3\times3} & 0_{3\times(N-3)} & G_u \end{bmatrix}, \quad \text{and} \quad d\mathbf{w} = \begin{bmatrix} d\mathbf{w}_v \\ d\mathbf{w}_m \\ d\mathbf{w}_u \end{bmatrix}. \tag{62}$$

It will be convenient to stack the columns of the matrices $B_{t|\mu,i}$ for $i = 0, \ldots, N$ in nine-component vectors $\mathbf{b}_i$, in the same way as equation (48) stacks the columns of A, and to define the $9\times N$ matrix

$$\mathcal{B} \equiv \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_N \end{bmatrix}. \tag{63}$$

This allows us to write the negative-log-likelihood function as

$$J = \tfrac{1}{2}\mathbf{x}^T F_{t|\mu}^x \mathbf{x} - \boldsymbol{\phi}_{t|\mu}^T \mathbf{x} - \mathbf{x}_A^T \mathbf{b}_0 - \mathbf{x}_A^T \mathcal{B}\mathbf{x} + \ldots, \tag{64}$$

with the partial derivatives with respect to the attitude vector

$$\frac{\partial J}{\partial \mathbf{x}_A} = -(\mathbf{b}_0 + \mathcal{B}\mathbf{x}) \tag{65}$$

and with respect to the bias vector

$$\frac{\partial J}{\partial \mathbf{x}} = -(\boldsymbol{\phi}_{t|\mu} + \mathcal{B}^T \mathbf{x}_A - F_{t|\mu}^x \mathbf{x}) = -[\boldsymbol{\phi}_{t|\mu} + \boldsymbol{\eta}_{\mu|\nu}(A) - F_{t|\mu}^x \mathbf{x}], \tag{66}$$

where $\boldsymbol{\eta}_{\mu|\nu}(A)$ is defined by equation (32). We are now prepared to evaluate the right side of equation (52) one term at a time.

$$\sum_{i=1}^{N+9} \frac{\partial f_i}{\partial \xi_i} = \sum_{i=1}^{9} \frac{\partial[\Gamma_A^T(\mathbf{x}_d - \boldsymbol{\omega}_g)]_i}{\partial \xi_i} = 0, \tag{67}$$

because ξ_i does not appear in the ith row of Γ_A. The next term is

$$-\sum_{i=1}^{N+9} f_i \frac{\partial J}{\partial \xi_i} = -\sum_{i=1}^{9} f_i \frac{\partial J}{\partial \xi_i} = (\mathbf{x}_d - \boldsymbol{\omega}_g)^T \Gamma_A(\mathbf{b}_0 + \mathcal{B}\mathbf{x}) = (\mathbf{x}_d - \boldsymbol{\omega}_g)^T \boldsymbol{\zeta}, \tag{68}$$

where

$$\boldsymbol{\zeta} \equiv \Gamma_A(\mathbf{b}_0 + \mathcal{B}\mathbf{x}). \tag{69}$$

Using the definitions of Γ_A, $\mathbf{b}_0$, $\mathcal{B}$, and $B_{\mu|\nu}(\mathbf{x})$ shows that $\boldsymbol{\zeta}$ has the components

$$\zeta_i = \operatorname{tr}\left\{B_{t|\mu}^T(\mathbf{x})[\mathbf{e}_i \times]A\right\} \quad \text{for} \quad i = 1, 2, 3. \tag{70}$$

10

The fact that the ith row of G does not depend on ξ_i allows us to write the vector $\mathbf{g}$ as

$$g_i = \sum_{j,k=1}^{N+9} G_{ik} \left(\frac{\partial G_{jk}}{\partial \xi_j} - G_{jk} \frac{\partial J}{\partial \xi_j} \right) = -\sum_{j=1}^{N+9} (GG^T)_{ij} \frac{\partial J}{\partial \xi_j}. \tag{71}$$

Writing

$$GG^T = \begin{bmatrix} \Gamma_A^T Q_v \Gamma_A & 0_{3\times(N-3)} & 0_{3\times 3} \\ 0_{(N-3)\times 3} & Q_m & 0_{(N-3)\times 3} \\ 0_{3\times 3} & 0_{3\times(N-3)} & Q_u \end{bmatrix} = \begin{bmatrix} \Gamma_A^T Q_v \Gamma_A & 0_{3\times N} \\ 0_{N\times 3} & Q_b \end{bmatrix}, \tag{72}$$

with the customary process noise power spectral density matrices defined by

$$Q_v \equiv G_v G_v^T, \quad Q_m \equiv G_m G_m^T, \quad \text{and} \quad Q_u \equiv G_u G_u^T, \tag{73}$$

and using equations (65), (66), and (69) gives

$$\mathbf{g} = \begin{bmatrix} \Gamma_A^T Q_v \zeta \\ Q_b [\phi_{t|\mu} + \eta_{\mu|v}(A) - F_{t|\mu}^x \mathbf{x}] \end{bmatrix}. \tag{74}$$

Then the last two terms on the right side of equation (52) are given by

$$\frac{1}{2} \sum_{i=1}^{N+9} g_i \frac{\partial J}{\partial \xi_i} = -\frac{1}{2} \left\{ \zeta^T Q_v \zeta + [\phi_{t|\mu} + \eta_{\mu|v}(A) - F_{t|\mu}^x \mathbf{x}]^T Q_b [\phi_{t|\mu} + \eta_{\mu|v}(A) - F_{t|\mu}^x \mathbf{x}] \right\} \tag{75}$$

and

$$-\frac{1}{2} \sum_{i=1}^{N+9} \frac{\partial g_i}{\partial \xi_i} = -\frac{1}{2} \sum_{i=1}^{9} \frac{\partial (\Gamma_A^T Q_v \zeta)_i}{\partial \xi_i} - \frac{1}{2} \sum_{i=1}^{N} \frac{\partial \{ Q_b [\phi_{t|\mu} + \eta_{\mu|v}(A) - F_{t|\mu}^x \mathbf{x}] \}_i}{\partial x_i}. \tag{76}$$

Because ξ_i does not appear in the ith row of Γ_A, the first term on the right side of equation (76) can be written as

$$-\frac{1}{2} \sum_{i=1}^{9} \sum_{j=1}^{3} (\Gamma_A^T Q_v)_{ij} (\partial \zeta_j / \partial \xi_i) = \frac{1}{2} \operatorname{tr}(\Gamma_A^T Q_v \Gamma_B) = \frac{1}{2} \operatorname{tr}(\Gamma_B \Gamma_A^T Q_v), \tag{77}$$

where differentiating equation (69) shows that Γ_B is constructed from the columns of $B_{t|\mu}(\mathbf{x})$ in the same way that equation (60) constructs Γ_A from the columns of A. Then matrix multiplication gives

$$\Gamma_B \Gamma_A^T = I_{3\times 3} \operatorname{tr}[AB_{t|\mu}^T(\mathbf{x})] - AB_{t|\mu}^T(\mathbf{x}). \tag{78}$$

The second term on the right side of equation (76) can also be simplified, so the right side is

$$-\frac{1}{2} \sum_{i=1}^{N+9} \frac{\partial g_i}{\partial \xi_i} = \frac{1}{2} \operatorname{tr}[B_{t|\mu}^T(\mathbf{x})(I_{3\times 3} \operatorname{tr} Q_v - Q_v)A] + \frac{1}{2} \operatorname{tr}(Q_b F_{t|v}^x). \tag{79}$$

Putting all these terms together and omitting terms that depend neither on A nor on $\mathbf{x}$ gives

$$\frac{1}{2} \mathbf{x}^T \dot{F}_{t|\mu}^x \mathbf{x} - \dot{\phi}_{t|\mu}^T \mathbf{x} - \operatorname{tr}(\dot{B}_{t|\mu,0}^T A) - \sum_{k=1}^{N} x_k \operatorname{tr}(\dot{B}_{t|\mu,k}^T A) = (\mathbf{x}_d - \omega_g)^T \zeta + \frac{1}{2} \operatorname{tr}[B_{t|\mu}^T(\mathbf{x})(I_{3\times 3} \operatorname{tr} Q_v - Q_v)A]$$

$$-\frac{1}{2} \left\{ \zeta^T Q_v \zeta + [\phi_{t|\mu} + \eta_{\mu|v}(A) - F_{t|\mu}^x \mathbf{x}]^T Q_b [\phi_{t|\mu} + \eta_{\mu|v}(A) - F_{t|\mu}^x \mathbf{x}] \right\} + \dots. \tag{80}$$

We must find approximations for three of the terms in the expansion of the right side of this equation because their functional dependence on A and $\mathbf{x}$ does not agree with any of the terms on the left side.

Equation (70) shows that the first term on the right side of equation (80) is quadratic in **x** and linear in A. We expand **x** about its expectation $\hat{\mathbf{x}}_{t|\mu}$ to write this term as

$$(\mathbf{x}_d - \boldsymbol{\omega}_g)^T \boldsymbol{\zeta} = \mathrm{tr}\left\{ AB_{t|\mu}^T(\mathbf{x})[(\mathbf{x}_d - \boldsymbol{\omega}_g)\times] \right\} = \mathrm{tr}\left\{ AB_{t|\mu}^T(\mathbf{x})[(\hat{\mathbf{x}}_d - \boldsymbol{\omega}_g)\times] \right\}$$
$$+ \mathrm{tr}\left\{ AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[(\mathbf{x}_d - \hat{\mathbf{x}}_d)\times] \right\} + \mathrm{tr}\left\{ A\sum_{j=1}^N [B_{t|\mu,j}^T(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu})_j][(\mathbf{x}_d - \hat{\mathbf{x}}_d)\times] \right\}, \quad (81)$$

omitting the subscript $t|\mu$ on $\hat{\mathbf{x}}_d$. We approximate A by $\hat{A}_{t|\mu}$ in the last term in this equation, introducing errors of order $\left| \mathbf{x} - \hat{\mathbf{x}}_{t|\mu} \right|^2 \left\| A - \hat{A}_{t|\mu} \right\|$. This gives the approximation to the first term on the right side of equation (80)

$$(\mathbf{x}_d - \boldsymbol{\omega}_g)^T \boldsymbol{\zeta} \approx \mathrm{tr}\left\{ AB_{t|\mu}^T(\mathbf{x})[(\hat{\mathbf{x}}_d - \boldsymbol{\omega}_g)\times] \right\} + \mathrm{tr}\left\{ AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[(\mathbf{x}_d - \hat{\mathbf{x}}_d)\times] \right\} + (\mathbf{x}_d - \hat{\mathbf{x}}_d)^T Z_{t|\mu}(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu}), \quad (82)$$

where $Z_{t|\mu}$ is the $3\times N$ matrix

$$(Z_{t|\mu})_{ij} = \mathrm{tr}\left([\mathbf{e}_i \times] \hat{A}_{t|\mu} B_{t|\mu,j}^T \right), \quad (83)$$

The term $\frac{1}{2}\boldsymbol{\zeta}^T Q_v \boldsymbol{\zeta}$ on the right side of equation (80) is quadratic in both **x** and A since $\boldsymbol{\zeta}$ is linear in both of these variables. We begin analyzing this term by writing $\boldsymbol{\zeta}$ as

$$\zeta_i = \mathrm{tr}\left\{ [\mathbf{e}_i \times]AB_{t|\mu}^T(\mathbf{x}) \right\} = \mathrm{tr}\left\{ [\mathbf{e}_i \times]AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) \right\} + \mathrm{tr}\left\{ [\mathbf{e}_i \times]A\sum_{j=1}^N [B_{t|\mu,j}^T(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu})_j] \right\}. \quad (84)$$

We approximate A by $\hat{A}_{t|\mu}$ in the last term in this equation and use equation (83) to give

$$\zeta_i \approx \mathrm{tr}\left\{ [\mathbf{e}_i \times]AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) \right\} + [Z_{t|\mu}(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu})]_i, \quad (85)$$

with errors of order $\left| \mathbf{x} - \hat{\mathbf{x}}_{t|\mu} \right| \left\| A - \hat{A}_{t|\mu} \right\|$. The symmetry of the matrix $\hat{A}_{t|\mu} B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) = U_+ S' U_+^T$ causes $\boldsymbol{\zeta}$ to vanish when $\mathbf{x} = \hat{\mathbf{x}}_{t|\mu}$ and $A = \hat{A}_{t|\mu}$, though, so the approximation

$$\frac{1}{2}\boldsymbol{\zeta}^T Q_v \boldsymbol{\zeta} \approx \frac{1}{2}\sum_{i,j=1}^3 \mathrm{tr}\left\{ [\mathbf{e}_i \times]AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) \right\}(Q_v)_{ij}\mathrm{tr}\left\{ [\mathbf{e}_j \times]AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) \right\}$$
$$+ \sum_{i=1}^3 \mathrm{tr}\left\{ [\mathbf{e}_i \times]AB_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) \right\}\left[Q_v Z_{t|\mu}(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu}) \right]_i + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu})^T Z_{t|\mu}^T Q_v Z_{t|\mu}(\mathbf{x} - \hat{\mathbf{x}}_{t|\mu}). \quad (86)$$

has errors on the same order of magnitude as the approximation in equation (82). The first term on the right side of equation (86) is quadratic in A, so we approximate it as in the measurement update by

$$\frac{1}{2}\sum_{i,j=1}^3 \mathrm{tr}\left\{ B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[\mathbf{e}_i \times]A \right\}(Q_v)_{ij}\mathrm{tr}\left\{ B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[\mathbf{e}_j \times]A \right\} \approx \psi_v + \mathrm{tr}(\Psi_v^T A). \quad (87)$$

For the MAP estimator, we only need to compute the $\ell = 1$ part Ψ_v, which is given by equation (39) as

$$(\Psi_v)_{mn} = \frac{3}{2}\sum_{i,j=1}^3 (Q_v)_{ij} \int_{\mathrm{SO(3)}} A_{mn}\mathrm{tr}\left\{ B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[\mathbf{e}_i \times]A \right\}\mathrm{tr}\left\{ B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[\mathbf{e}_j \times]A \right\}dA$$
$$= \frac{1}{4}\sum_{i,j,k,\ell,p,q=1}^3 (Q_v)_{ij}\left\{ [\mathbf{e}_i \times] B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) \right\}_{k\ell}\left\{ [\mathbf{e}_j \times] B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) \right\}_{pq} \varepsilon_{mkp}\varepsilon_{n\ell q}. \quad (88)$$

The term $\frac{1}{2}\boldsymbol{\eta}_{\mu|v}^T(A)Q_b\,\boldsymbol{\eta}_{\mu|v}(A)$ on the right side of equation (80) is quadratic in A, so we approximate it as

$$\frac{1}{2}\boldsymbol{\eta}_{\mu|v}^T(A)Q_b\,\boldsymbol{\eta}_{\mu|v}(A) = \frac{1}{2}\sum_{p,q=1}^{N} \mathrm{tr}(B_{t|\mu,p}^T A)(Q_b)_{pq}\,\mathrm{tr}(B_{t|\mu,q}^T A) \approx \psi_b + \mathrm{tr}(\Psi_b^T A), \tag{89}$$

In parallel with equation (88), we only need to compute the $\ell = 1$ part, which is given by

$$(\Psi_b)_{mn} = \frac{3}{2}\int_{SO(3)} A_{mn}\boldsymbol{\eta}_{\mu|v}^T(A)Q_b\,\boldsymbol{\eta}_{\mu|v}(A)\,dA = \frac{1}{4}\sum_{p,q=1}^{N}(Q_b)_{pq}\sum_{i,j,k,\ell=1}^{3}(B_{t|\mu,p})_{ij}(B_{t|\mu,q})_{k\ell}\varepsilon_{mik}\varepsilon_{nj\ell}. \tag{90}$$

Equating terms with the same dependence on A and $\mathbf{x}$ on the two sides of equation (80) with the above approximations for the first three terms on the right gives the final propagation equations for the parameters in the PDF, assuming that the components of $\mathbf{x}_d$ are the last three components in the bias vector $\mathbf{x}$:

$$\dot{F}_{t|\mu}^x = \begin{bmatrix}0_{(N-3)\times N}\\ Z_{t|\mu}\end{bmatrix} + \begin{bmatrix}0_{N\times(N-3)} & Z_{t|\mu}^T\end{bmatrix} - F_{t|\mu}^x Q_b F_{t|\mu}^x - Z_{t|\mu}^T Q_v Z_{t|\mu} \tag{91}$$

$$\dot{\boldsymbol{\phi}}_{t|\mu} = \begin{bmatrix}0_{(N-3)\times N}\\ Z_{t|\mu}\end{bmatrix}\hat{\mathbf{x}}_{t|\mu} + Z_{t|\mu}^T(\hat{\mathbf{x}}_d - Q_v Z_{t|\mu}\hat{\mathbf{x}}_{t|\mu}) - F_{t|\mu}^x Q_b \boldsymbol{\phi}_{t|\mu} \tag{92}$$

$$\dot{B}_{t|\mu,0} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]B_{t|\mu,0} - \frac{1}{2}(I_{3\times3}\,\mathrm{tr}Q_v - Q_v)B_{t|\mu,0} + \sum_{i=1}^{N}(Q_b\boldsymbol{\phi}_{t|\mu})_i B_{t|\mu,i}$$
$$-[(\hat{\mathbf{x}}_d - Q_v Z_{t|\mu}\hat{\mathbf{x}}_{t|\mu})\times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) + \Psi_v + \Psi_b \tag{93}$$

$$\dot{B}_{t|\mu,i} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]B_{t|\mu,i} - \frac{1}{2}(I_{3\times3}\,\mathrm{tr}Q_v - Q_v)B_{t|\mu,i} - \sum_{j=1}^{N}(F_{t|\mu}^x Q_b)_{ij} B_{t|\mu,j}$$
$$-[(Q_v \mathbf{z}_{t|\mu,i})\times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) \qquad\qquad \text{for} \quad 1 \le i \le N-3 \tag{94a}$$

$$\dot{B}_{t|\mu,i} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]B_{t|\mu,i} - \frac{1}{2}(I_{3\times3}\,\mathrm{tr}Q_v - Q_v)B_{t|\mu,i} - \sum_{j=1}^{N}(F_{t|\mu}^x Q_b)_{ij} B_{t|\mu,j}$$
$$-[(Q_v \mathbf{z}_{t|\mu,i} - \mathbf{e}_{i-(N-3)})\times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) \qquad \text{for} \quad i = N-2, N-1, N, \tag{94b}$$

where $\mathbf{z}_{t|\mu,i}$ is the ith column of $Z_{t|\mu}$. The first two terms on the right side of equation (93) give the propagation of Filter QUEST with a fading memory [31], and equation (94) is similar. The interesting new terms in the orthogonal filter are the nonlinear coupling terms in these equations.

AN ASIDE

Aside from the approximations we made, one objection can be raised to the derivation of the Fokker-Planck equation. Computing partial derivatives with respect to $\mathbf{x}_A$ implicitly assumes that these parameters can be varied independently, but independent variations of these parameters takes A out of SO(3). To assuage these fears, the derivation was repeated with the quaternion ray representation of the attitude [11]

$$A(\mathbf{q}, q_4) = \left(|\mathbf{q}|^2 + q_4^2\right)^{-1}\left\{(q_4^2 - |\mathbf{q}|^2)I_{3\times3} + 2\mathbf{q}\mathbf{q}^T - 2q_4[\mathbf{q}\times]\right\}. \tag{95}$$

which guarantees that A is an orthogonal matrix for any independent variations of $\mathbf{q}$ and q_4, and with the $(N+4)$-component state vector

$$\boldsymbol{\xi} \equiv \begin{bmatrix}\mathbf{q}\\ q_4\\ \mathbf{x}\end{bmatrix}. \tag{96}$$

This derivation also yielded equation (80), with considerably more effort.

13

COVARIANCE

None of the computations in the orthogonal filter requires a covariance matrix. This is a disadvantage in practice, since the orthogonal filter provides no natural way to perform a data-independent covariance analysis, as is straightforward in an EKF. We can, however, compute the Cramér-Rao lower bound on the covariance from the Fisher information matrix [1].

The most useful form of the covariance matrix is in terms of the covariance of a vector $\boldsymbol{\theta}$ of infinitesimal rotation errors and the vector $\tilde{\mathbf{x}} \equiv \mathbf{x} - \hat{\mathbf{x}}_{\mu|\nu}$ of bias errors. We write the attitude matrix as

$$A = \exp(-[\boldsymbol{\theta}\times])A_{true} \approx (I_{3\times3} - [\boldsymbol{\theta}\times] + \tfrac{1}{2}[\boldsymbol{\theta}\times]^2)\hat{A}_{\mu|\nu},\qquad(97)$$

expanding the exponential and approximating the true attitude by its estimate. Substituting this into equation (27), using equation (28), and expanding gives

$$\begin{aligned} J_{A_\mu,\mathbf{x}_\mu|Y_\nu}(A,\mathbf{x}) = &\tfrac{1}{2}\tilde{\mathbf{x}}^T F_{\mu|\nu}^x\tilde{\mathbf{x}} + \hat{\mathbf{x}}_{\mu|\nu}^T F_{\mu|\nu}^x\tilde{\mathbf{x}} - \boldsymbol{\phi}_{\mu|\nu}^T\tilde{\mathbf{x}} + \operatorname{tr}\{B_{\mu|\nu}^T(\hat{\mathbf{x}}_{\mu|\nu})[\boldsymbol{\theta}\times]\hat{A}_{\mu|\nu}\} \\ &- \sum_{j=1}^N \tilde{x}_j\operatorname{tr}(B_{\mu|\nu,j}^T\hat{A}_{\mu|\nu}) + \sum_{j=1}^N \tilde{x}_j\operatorname{tr}\{B_{\mu|\nu,j}^T[\boldsymbol{\theta}\times]\hat{A}_{\mu|\nu}\} - \tfrac{1}{2}\operatorname{tr}\{B_{\mu|\nu}^T(\hat{\mathbf{x}}_{\mu|\nu})[\boldsymbol{\theta}\times]^2\hat{A}_{\mu|\nu}\}+\dots, \end{aligned}\qquad(98)$$

where constant terms and terms of higher than second order have been omitted. Evaluating the second-order partial derivatives of this negative-log-likelihood function gives the Fisher information matrix

$$F_{\mu|\nu} \equiv E\left\{\begin{bmatrix} \partial^2 J/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta} & \partial^2 J/\partial\boldsymbol{\theta}\,\partial\tilde{\mathbf{x}} \\ \partial^2 J/\partial\tilde{\mathbf{x}}\,\partial\boldsymbol{\theta} & \partial^2 J/\partial\tilde{\mathbf{x}}\,\partial\tilde{\mathbf{x}} \end{bmatrix}\right\} = \begin{bmatrix} F_{\mu|\nu}^\theta & Z_{\mu|\nu} \\ Z_{\mu|\nu}^T & F_{\mu|\nu}^x \end{bmatrix},\qquad(99)$$

where $Z_{\mu|\nu}$ is the matrix defined by equation (83). The attitude information matrix $F_{\mu|\nu}^\theta$ is given by

$$F_{\mu|\nu}^\theta = I_{3\times3}\operatorname{tr}[\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})] - \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})\qquad(100)$$

where the SVD has been used to obtain the second equality. The partitions of the covariance matrix

$$P_{\mu|\nu} \equiv F_{\mu|\nu}^{-1} \equiv \begin{bmatrix} P_{\mu|\nu}^\theta & P_{\mu|\nu}^c \\ (P_{\mu|\nu}^c)^T & P_{\mu|\nu}^x \end{bmatrix},\qquad(101)$$

are computed using the standard formulas for the inversion of a partitioned matrix [35];

$$P_{\mu|\nu}^\theta = \{I_{3\times3}\operatorname{tr}[\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})] - \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) - Z_{\mu|\nu}(F_{\mu|\nu}^x)^{-1}Z_{\mu|\nu}^T\}^{-1}\qquad(102)$$

$$P_{\mu|\nu}^x = \left[\!\left[F_{\mu|\nu}^x - Z_{\mu|\nu}^T\{I_{3\times3}\operatorname{tr}[\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})] - \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})\}^{-1}Z_{\mu|\nu}\right]\!\right]^{-1}\qquad(103)$$

$$P_{\mu|\nu}^c = -P_{\mu|\nu}^\theta Z_{\mu|\nu}(F_{\mu|\nu}^x)^{-1} = -\{I_{3\times3}\operatorname{tr}[\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})] - \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})\}^{-1}Z_{\mu|\nu}P_{\mu|\nu}^x.\qquad(104)$$

The information matrix and covariance matrix only depend on the skew parts of the matrices $\hat{A}_{\mu|\nu}B_{\mu|\nu,j}^T$, so only on three independent linear combinations of the elements of $B_{\mu|\nu,j}$. This shows that the covariance does not contain all the information in $B_{\mu|\nu,j}$, as we noted earlier. It is impossible to know which linear combination of the components of $B_{\mu|\nu,j}$ appears in the covariance without knowing $\hat{A}_{\mu|\nu}$, however.

With the dynamic model of equation (55) and (57), the (N+3)-component state vector [7]

$$\boldsymbol{\xi} \equiv \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x}_m - \hat{\mathbf{x}}_m \\ \mathbf{x}_d - \hat{\mathbf{x}}_d \end{bmatrix} , \tag{105}$$

obeys the linearized SDE

$$d\boldsymbol{\xi} = \mathcal{F}\boldsymbol{\xi}\,dt + \mathcal{G}\,d\mathbf{w} , \tag{106}$$

with

$$\mathcal{F} = \begin{bmatrix} -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times] & 0_{3\times(N-3)} & -I_{3\times3} \\ 0_{(N-3)\times3} & 0_{(N-3)\times(N-3)} & 0_{(N-3)\times3} \\ 0_{3\times3} & 0_{3\times(N-3)} & 0_{3\times3} \end{bmatrix} \quad \text{and} \quad \mathcal{G} = \begin{bmatrix} G_v & 0_{3\times(N-3)} & 0_{3\times3} \\ 0_{(N-3)\times3} & G_m & 0_{(N-3)\times3} \\ 0_{3\times3} & 0_{3\times(N-3)} & G_u \end{bmatrix} . \tag{107}$$

The covariance of this state vector $\boldsymbol{\xi}$ obeys the propagation equation

$$\dot{P}_{t|\mu} = \mathcal{F}P_{t|\mu} + P_{t|\mu}\mathcal{F}^T + \mathcal{G}\mathcal{G}^T , \tag{108}$$

so the Fisher information matrix, the inverse of the covariance, obeys

$$\dot{F}_{t|\mu} = -F_{t|\mu}\mathcal{F} - \mathcal{F}^T F_{t|\mu} - F_{t|\mu}\mathcal{G}\mathcal{G}^T F_{t|\mu} . \tag{109}$$

Then the partitions of the information matrix propagate as

$$\dot{F}_{t|\mu}^\theta = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]F_{t|\mu}^\theta - F_{t|\mu}^\theta[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]^T - F_{t|\mu}^\theta Q_v F_{t|\mu}^\theta - Z_{t|\mu}Q_b Z_{t|\mu}^T . \tag{110}$$

$$\dot{Z}_{t|\mu} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]Z_{t|\mu} + \begin{bmatrix} 0_{3\times(N-3)} & F_{t|\mu}^\theta \end{bmatrix} - F_{t|\mu}^\theta Q_v Z_{t|\mu} - Z_{t|\mu}Q_b F_{t|\mu}^x . \tag{111}$$

$$\dot{F}_{t|\mu}^x = \begin{bmatrix} 0_{(N-3)\times N} \\ Z_{t|\mu} \end{bmatrix} + \begin{bmatrix} 0_{N\times(N-3)} & Z_{t|\mu}^T \end{bmatrix} - F_{t|\mu}^x Q_b F_{t|\mu}^x - Z_{t|\mu}^T Q_v Z_{t|\mu} . \tag{112}$$

It is reassuring that equation (112) is identical to equation (91) of the orthogonal filter. It is also interesting that propagation moves the information down from $F_{t|\mu}^\theta$ to $Z_{t|\mu}$ and then from $Z_{t|\mu}$ to $F_{t|\mu}^x$. This is exactly opposite to the covariance propagation, in which the errors propagate up from $P_{t|\mu}^x$ to $P_{t|\mu}^c$ and then from $P_{t|\mu}^c$ to $P_{t|\mu}^\theta$. This counterintuitive behavior can be explained by the observation that knowledge of the attitude time dependence provides information about the biases.

PROPAGATION OF THE MAP ESTIMATES

Since the MAP estimation of the attitude is an orthogonal matrix, its derivative must be given by

$$\dot{\hat{A}}_{t|\mu} = -[\hat{\boldsymbol{\omega}}_{t|\mu} \times]\hat{A}_{t|\mu} \tag{113}$$

for some angular velocity vector $\hat{\boldsymbol{\omega}}_{t|\mu}$ that is to be determined. Substituting equation (17) gives

$$[\hat{\boldsymbol{\omega}}_{t|\mu} \times] = -\dot{\hat{A}}_{t|\mu}\hat{A}_{t|\mu}^T = -(\dot{U}_+ V_+^T + U_+ \dot{V}_+^T)V_+ U_+^T = U_+([\boldsymbol{\omega}_U \times] + [\boldsymbol{\omega}_V \times])U_+^T , \tag{114}$$

where

$$[\boldsymbol{\omega}_U \times] \equiv -U_+^T \dot{U}_+ \quad \text{and} \quad [\boldsymbol{\omega}_V \times] \equiv -\dot{V}_+^T V_+ . \tag{115}$$

This is equivalent to the vector equation

$$\hat{\boldsymbol{\omega}}_{t|\mu} = U_+(\boldsymbol{\omega}_U + \boldsymbol{\omega}_V) . \tag{116}$$

15

Differentiating equation (15) and using equation (115) gives

$$\dot{B}_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) = U_+(-[\boldsymbol{\omega}_U \times]S' + \dot{S}' - S'[\boldsymbol{\omega}_V \times])V_+^T. \tag{117}$$

If we now define the vector $\boldsymbol{\beta}$ by

$$[\boldsymbol{\beta} \times] \equiv \hat{A}_{t|\mu} \dot{B}_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}) - \dot{B}_{t|\mu}(\hat{\mathbf{x}}_{t|\mu})\hat{A}_{t|\mu}^T, \tag{118}$$

we find after a little algebra that

$$[U_+^T\boldsymbol{\beta} \times] = S'[(\boldsymbol{\omega}_U + \boldsymbol{\omega}_V) \times] + [(\boldsymbol{\omega}_U + \boldsymbol{\omega}_V) \times]S' = S'[U_+^T\hat{\boldsymbol{\omega}}_{t|\mu} \times] + [U_+^T\hat{\boldsymbol{\omega}}_{t|\mu} \times]S'. \tag{119}$$

But this means that

$$U_+^T\boldsymbol{\beta} = (I_{3\times3} \operatorname{tr}S' - S')U_+^T\hat{\boldsymbol{\omega}}_{t|\mu} \tag{120}$$

or

$$\boldsymbol{\beta} = \{I_{3\times3} \operatorname{tr}[\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})] - \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})\}\hat{\boldsymbol{\omega}}_{t|\mu}. \tag{121}$$

In order to compute $\boldsymbol{\beta}$ from equation (118), we need to evaluate

$$\dot{B}_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) = \dot{B}_{t|\mu,0} + \sum_{k=1}^{N} \dot{B}_{t|\mu,k}\hat{x}_{t|\mu,k} + \sum_{k=1}^{N} B_{t|\mu,k}\dot{\hat{x}}_{t|\mu,k}. \tag{122}$$

The first two terms are given by equations (93) and (94), and differentiating equation (34) gives

$$\dot{\hat{\mathbf{x}}}_{t|\mu} = (F_{t|\mu}^x)^{-1}\dot{\boldsymbol{\eta}}_{t|\mu} + Q_b\boldsymbol{\eta}_{t|\mu}. \tag{123}$$

after substitutions from equations (91) and (92). Putting all this together gives, after several cancellations,

$$\dot{B}_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) = \left\{-[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d) \times] - \tfrac{1}{2}(I_{3\times3} \operatorname{tr}Q_v - Q_v)\right\}B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) + \sum_{i=1}^{N}[(F_{t|\mu}^x)^{-1}\dot{\boldsymbol{\eta}}_{t|\mu}]_i B_{t|\mu,i} + \Psi_v + \Psi_b. \tag{124}$$

We now make two simplifying assumptions. We first assume that the gyro drift rate process noise is isotropic, i.e. that $Q_v = \sigma_v^2 I_{3\times3}$, so that

$$\tfrac{1}{2}(I_{3\times3} \operatorname{tr}Q_v - Q_v) = \sigma_v^2 I_{3\times3} = Q_v$$

and

$$\Psi_v = \tfrac{1}{2}\sigma_v^2 \operatorname{adj}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu}). \tag{125}$$

where adj(.) denotes the classical adjoint [35]. We also assume that Q_b is diagonal, so that

$$\Psi_b = \tfrac{1}{2}\sum_{i=1}^{N}(Q_b)_{ii} \operatorname{adj}B_{t|\mu,i}^T. \tag{126}$$

With these simplifications, substituting equation (124) into equation (118) gives

$$[\boldsymbol{\beta} \times] = \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d) \times] + [(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d) \times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu})\hat{A}_{t|\mu}^T$$
$$+ \sum_{i=1}^{N}[(F_{t|\mu}^x)^{-1}\dot{\boldsymbol{\eta}}_{t|\mu}]_i(\hat{A}_{t|\mu}B_{t|\mu,i}^T - B_{t|\mu,i}\hat{A}_{t|\mu}^T) + \tfrac{1}{2}\sum_{i=1}^{N}(Q_b)_{ii}[\operatorname{adj}(B_{t|\mu,i}\hat{A}_{t|\mu}^T) - \operatorname{adj}(\hat{A}_{t|\mu}B_{t|\mu,i}^T)]. \tag{127}$$

The $\tfrac{1}{2}(I_{3\times3} \operatorname{tr}Q_v - Q_v)B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu})$ and Ψ_v terms in equation (124) give no contribution owing to the symmetry of $\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})$. Equation (127) is equivalent to the vector equation

$$\boldsymbol{\beta} = \{I_{3\times3} \operatorname{tr}[\hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})] - \hat{A}_{t|\mu}B_{t|\mu}^T(\hat{\mathbf{x}}_{t|\mu})\}(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d) - Z_{t|\mu}(F_{t|\mu}^x)^{-1}\dot{\boldsymbol{\eta}}_{t|\mu} - \tfrac{1}{2}\sum_{i=1}^{N}(Q_b)_{ii} B_{t|\mu,i}\hat{A}_{t|\mu}^T\mathbf{z}_{t|\mu,i}. \tag{128}$$

16

We find the time derivative of $\eta_{t|\mu}$ from its definition in equation (32). With our simplifying assumptions,

$$(\dot{\eta}_{t|\mu})_i = \text{tr}(\dot{B}^T_{t|\mu,i}\hat{A}_{t|\mu} + B^T_{t|\mu,i}\dot{\hat{A}}_{t|\mu}) = \text{tr}\{[(\hat{\omega}_{t|\mu} - \omega_g + \hat{x}_d) \times]B^T_{t|\mu,i}\hat{A}_{t|\mu}\} - (\sigma_v^2 \eta_{t|\mu} + F^x Q_b \eta_{t|\mu})_i , \quad (129)$$

using the fact that the trace of the product of a skew-symmetric matrix and a symmetric matrix vanishes. This gives

$$\dot{\eta}_{t|\mu} = -Z^T_{t|\mu}(\hat{\omega}_{t|\mu} - \omega_g + \hat{x}_d) - \sigma_v^2 \eta_{t|\mu} - F^x Q_b \eta_{t|\mu} . \quad (130)$$

Substituting into equation (121) then yields

$$\hat{\omega}_{t|\mu} = \omega_g - \hat{x}_d + \Delta\omega_{t|\mu} , \quad (131)$$

with

$$\Delta\omega_{t|\mu} \equiv P^\theta_{t|\mu}\{Z_{t|\mu}[\sigma_v^2(F^x)^{-1} + Q_b]\eta_{t|\mu} - \frac{1}{2}\sum_{i=1}^N (Q_b)_{ii} B_{t|\mu,i}\hat{A}^T_{t|\mu}z_{t|\mu,i}\} , \quad (132)$$

where $P^\theta_{\mu|v}$ is given by equation (102). Substituting these results back into equation (123) gives

$$F^x_{t|\mu}\dot{\hat{x}}_{t|\mu} = -Z^T_{t|\mu}\Delta\omega_{t|\mu} - \sigma_v^2 \eta_{t|\mu} . \quad (133)$$

We see that the orthogonal filter includes effects that are ignored by the EKF, which has $\hat{\omega}^{EKF}_{t|\mu} = \omega_g - \hat{x}_d$ and $\dot{\hat{x}}^{EKF}_{t|\mu} = 0_N$ with the dynamic model of equations (50) and (62). It is worth noting that the extra term in equation (131) bears no resemblance to the extra term found in a second-order Kalman filter [7].

ALGORITHM SUMMARY

The equations of the orthogonal attitude filter can be simply summarized. The parameters $F^x_{t|\mu}$, $\phi_{t|\mu}$, and $B_{t|\mu,i}$ specify the PDF. A priori values of $F^x_{0|0}$ and $\hat{x}_{0|0}$ are needed to initialize the filter. Then

$$\phi_{0|0} \equiv F^x_{0|0}\hat{x}_{0|0} \quad (134)$$

and

$$B_{0|0,i} \quad \text{for} \quad i = 1, \ldots, N . \quad (135)$$

If there is no a priori attitude estimate, $B_{0|0,0}$ is also initialized to zero. If there is an initial attitude estimate $\hat{A}_{0|0}$ with inverse covariance $F^\theta_{0|0}$, $B_{0|0,0}$ is initialized to [30]

$$B_{0|0,0} = \left[\frac{1}{2}(\text{tr} F^\theta_{0|0})I_{3\times3} - F^\theta_{0|0}\right]\hat{A}_{0|0} . \quad (136)$$

MAP attitude and bias estimates are found by iteratively solving the following equations:

$$USV^T = B_{\mu|v,0} + \sum_{k=1}^N B_{\mu|v,k}\hat{x}_k , \quad (137)$$

$$\hat{A}_{\mu|v} = U \text{diag}([1, 1, \det U \det V])V^T , \quad (138)$$

$$[\eta_{\mu|v}(\hat{A}_{\mu|v})]_k = \text{tr}(B^T_{\mu|v,k}\hat{A}_{\mu|v}) \quad \text{for} \quad k = 1, \ldots, N , \quad (139)$$

$$\hat{x}_{\mu|v} = (F^x_{\mu|v})^{-1}[\phi_{\mu|v} + \eta_{\mu|v}(\hat{A}_{\mu|v})] . \quad (140)$$

For the case of attitude determination using a triaxial magnetometer, $\mathbf{x}_m$ is a vector of magnetometer biases (assumed to be in the first three components of $\mathbf{x}$), and measurement information is incorporated by

$$F_{\mu|\mu}^x = F_{\mu|\mu-1}^x + \begin{bmatrix} F_\mu^m & 0_{3\times(N-3)} \\ 0_{(N-3)\times 3} & 0_{(N-3)\times(N-3)} \end{bmatrix} \tag{141}$$

$$\boldsymbol{\phi}_{\mu|\mu} = \boldsymbol{\phi}_{\mu|\mu-1} + \begin{bmatrix} F_\mu^m \mathbf{y}_\mu \\ \mathbf{0}_{N-3} \end{bmatrix} \tag{142}$$

$$B_{\mu|\mu,0} = B_{\mu|\mu-1,0} + F_\mu^m \mathbf{y}_\mu \mathbf{H}_\mu^T . \tag{143}$$

$$B_{\mu|\mu,i} = B_{\mu|\mu-1,i} - F_\mu^m \mathbf{e}_i \mathbf{H}_\mu^T . \qquad\qquad \text{for } i = 1, 2, 3 \tag{144a}$$

$$B_{\mu|\mu,i} = B_{\mu|\mu-1,i} \qquad\qquad \text{for } i > 3, \tag{144b}$$

where $\mathbf{y}_\mu$ is the measurement of the ambient magnetic field, F_μ^m is the inverse covariance of the assumed Gaussian measurement errors, $\mathbf{H}_\mu$ is the magnetic field in the coordinate frame to which the attitude is referenced, and the coordinate axis unit vectors $\mathbf{e}_i$ are defined by equation (47).

For attitude propagation using the gyro-sensed attitude rate vector $\boldsymbol{\omega}_g$ the body frame, $\mathbf{x}_d$ is a vector of gyro drift biases in the body frame (assumed to be in the last three components of $\mathbf{x}$), and the parameters in the PDF are propagated between measurements by

$$\dot{F}_{t|\mu}^x = \begin{bmatrix} 0_{(N-3)\times N} \\ Z_{t|\mu} \end{bmatrix} + \begin{bmatrix} 0_{N\times(N-3)} & Z_{t|\mu}^T \end{bmatrix} - F_{t|\mu}^x Q_b F_{t|\mu}^x - Z_{t|\mu}^T Q_v Z_{t|\mu} \tag{145}$$

$$\dot{\boldsymbol{\phi}}_{t|\mu} = \begin{bmatrix} 0_{(N-3)\times N} \\ Z_{t|\mu} \end{bmatrix} \hat{\mathbf{x}}_{t|\mu} + Z_{t|\mu}^T (\hat{\mathbf{x}}_d - Q_v Z_{t|\mu} \hat{\mathbf{x}}_{t|\mu}) - F_{t|\mu}^x Q_b \boldsymbol{\phi}_{t|\mu} \tag{146}$$

$$\dot{B}_{t|\mu,0} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]B_{t|\mu,0} - \tfrac{1}{2}(I_{3\times3}\,\mathrm{tr}Q_v - Q_v)B_{t|\mu,0} + \sum_{i=1}^N (Q_b\boldsymbol{\phi}_{t|\mu})_i B_{t|\mu,i}$$
$$-[(\hat{\mathbf{x}}_d - Q_v Z_{t|\mu}\hat{\mathbf{x}}_{t|\mu})\times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) + \Psi_v + \Psi_b \tag{147}$$

$$\dot{B}_{t|\mu,i} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]B_{t|\mu,i} - \tfrac{1}{2}(I_{3\times3}\,\mathrm{tr}Q_v - Q_v)B_{t|\mu,i} - \sum_{j=1}^N (F_{t|\mu}^x Q_b)_{ij} B_{t|\mu,j}$$
$$-[(Q_v\mathbf{z}_{t|\mu,i})\times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) \qquad\qquad \text{for } 1 \le i \le N-3 \tag{148a}$$

$$\dot{B}_{t|\mu,i} = -[(\boldsymbol{\omega}_g - \hat{\mathbf{x}}_d)\times]B_{t|\mu,i} - \tfrac{1}{2}(I_{3\times3}\,\mathrm{tr}Q_v - Q_v)B_{t|\mu,i} - \sum_{j=1}^N (F_{t|\mu}^x Q_b)_{ij} B_{t|\mu,j}$$
$$-[(Q_v\mathbf{z}_{t|\mu,i} - \mathbf{e}_{i-(N-3)})\times]B_{t|\mu}(\hat{\mathbf{x}}_{t|\mu}) \qquad\qquad \text{for } i = N-2, N-1, N, \tag{148b}$$

where Q_v and Q_b are process noise power spectral density matrices defined by equations (72) and (73),

$$(Z_{t|\mu})_{ij} = (\mathbf{z}_{t|\mu,j})_i = \mathrm{tr}\left([\mathbf{e}_i\times]\hat{A}_{t|\mu}B_{t|\mu,j}^T\right), \tag{149}$$

and where Ψ_v and Ψ_b are given by equations (88) and (90) in the general case, or by equations (125) and (126) with some simplifying assumptions about the process noise.

18

NUMERICAL EXAMPLE

We test this algorithm for a spacecraft in a circular orbit about the Earth with inclination 35 deg and radius 6775.19 km. All orbit perturbations are ignored, giving purely Keplerian motion. The spacecraft is assumed to maintain its body x (roll) axis along the velocity vector, its body z (yaw) axis along the nadir vector, and its body y (pitch) axis along the negative orbit normal. Then the orbit period is 92.5 minutes and the spacecraft angular velocity in the body frame is exactly

$$\boldsymbol{\omega} = -(2\pi/5550)\,\mathbf{e}_2 \text{ rad/sec.} \tag{150}$$

These parameters approximate those of NASA's Tropical Rainfall Measuring Mission (TRMM) spacecraft [43]. The Earth's magnetic field is modeled by a tilted dipole [44], which for our circular orbit gives

$$\mathbf{H}(\mathbf{r}) = H_0[3(\mathbf{m}\cdot\mathbf{r}_{unit})\mathbf{r}_{unit} - \mathbf{m}], \tag{151}$$

where H_0 is 25.54 μT, $\mathbf{r}_{unit}$ is a unit vector pointing from the center of the Earth to the spacecraft, and

$$\mathbf{m} = \begin{bmatrix} \sin\theta_m \sin\alpha_m \\ \sin\theta_m \cos\alpha_m \\ \cos\theta_m \end{bmatrix} \tag{152}$$

gives the direction of the Earth's dipole, with $\theta_m = 168.6$ deg and $\alpha_m = \alpha_0 + t \times 4.178 \times 10^{-3}$ deg/sec. These orbit, attitude, and magnetic field models are very crude, but they suffice for our purposes.

Magnetometer measurements are simulated with a bias of $+0.5$ μT and zero-mean white noise with a standard deviation of $\sigma_m = 0.5$ μT on each axis. Actual magnetic field errors have systematic components [45], but these are not relevant to the present filter comparisons. The filter uses $F_\mu^m = \sigma_m^{-2}I_{3\times3}$ and $Q_m = 0_{3\times3}$, since the magnetometer biases are assumed to be constant. The first measurement was simulated when the spacecraft crossed the ascending node, when the angle between the spacecraft x axis and the equatorial plane is 35 deg, the orbit inclination.

The spacecraft is assumed to have three-axis rate-integrating gyros that output an incremental angle vector $\boldsymbol{\theta}_{t|\mu}^{RIG}$. The RIGs are modeled with initial biases of $+0.1$ deg/hr, drift rate noise of $\sigma_v = 0.3$ μrad/sec$^{1/2}$, and drift rate ramp noise of $\sigma_u = 3\times10^{-4}$ μrad/sec$^{3/2}$ on each axis, which are like the TRMM gyros. The filter uses $Q_v = \sigma_v^2 I_{3\times3}$ and $Q_u = \sigma_u^2 I_{3\times3}$. The true gyro drift rates are simulated at discrete times by

$$\mathbf{x}_d(t) = \mathbf{x}_d(t_\mu) + \sigma_u(t - t_\mu)^{1/2}\mathbf{n}_u, \tag{153}$$

where the components of $\mathbf{n}_u$ are independent zero-mean normally distributed random variables with unity standard deviation. The true gyro incremental angle outputs are simulated by

$$\boldsymbol{\theta}_{t|\mu}^{RIG} = \left\{\boldsymbol{\omega} + \tfrac{1}{2}[\mathbf{x}_d(t) + \mathbf{x}_d(t_\mu)]\right\}(t - t_\mu) + \left[\sigma_v^2(t - t_\mu) + \tfrac{1}{12}\sigma_u^2(t - t_\mu)^3\right]^{1/2}\mathbf{n}_v, \tag{154}$$

where $\boldsymbol{\omega}$ is given by equation (150) and the components of $\mathbf{n}_v$ are also independent zero-mean normally distributed random variables with unity standard deviation. Equations (153) and (154) give biases and incremental angle outputs with variances and correlations that are consistent with equations (57) and (58). The gyros are assumed to be read out at the measurement processing rate, so we assume the constant rate

$$\boldsymbol{\omega}_g = (t_{\mu+1} - t_\mu)^{-1}\boldsymbol{\theta}_{\mu+1|\mu}^{RIG}, \tag{155}$$

to propagate the PDF between the measurements at times t_μ and $t_{\mu+1}$.

The simulation was performed using MATLAB® and its differential equation integrator ODE23. The initial attitude estimate at the time of the first measurement had the z axis along the nadir vector and the x axis in the equatorial plane, which means that the initial attitude estimate has a 35 deg yaw error. The initial estimates of the gyro drift bias and magnetometer bias were set to zero. The PDF was initialized with information equivalent to standard deviations on each axis of 40.5 deg in attitude, 0.1 deg/hour in the IRU drift biases, and 0.5 μT in the TAM biases. Iterations of equations (137)–(140) were terminated when

$$\Delta \hat{\mathbf{x}}^T F_{\mu|\nu}^x \Delta \hat{\mathbf{x}} < 10^{-8} \tag{156}$$

where $\Delta \hat{\mathbf{x}}$ is the change in the estimate of the bias state in successive iterations. Using tighter tolerances on the iteration convergence or different numerical integrators had no significant effect on the results.

Figure 1 shows the estimation results using the full orthogonal filter described by equations (134)–(149). Figure 2 shows the estimation results using the same filter except that the last term is deleted from the right side of equation (147). The results shown in Figures 1 and 2 are very close for the first 150 minutes of the simulation; the attitude estimates agree to within 0.01 deg, the IRU drift bias estimates to within 0.01 deg/hour, and the TAM bias estimates to within 1 nT. The pitch attitude estimates shown in Figure 1 are more accurate than those in Figure 2 at later times, but the estimator with Ψ_b included becomes unstable in roll and yaw. This behavior probably results from the approximations required to derive equations (91)–(94) from equation (80). It is possible that retaining the $\ell = 2$ contribution neglected by equation (89) would stabilize the orthogonal filter, at the expense of increased complexity. Further investigation shows that only the skew-symmetric part of the product $\hat{A}_{t|\mu}\Psi_b^T$ causes the divergence. The approximation in equation (87) turns out to be harmless because $\hat{A}_{t|\mu}\Psi_v^T$ is symmetric, as was pointed out below equation (127).

The same simulated data were used in an EKF employing the same measurement and process noise assumptions and the same initial conditions. The EKF, like the orthogonal filter, converges to useful estimates in about one orbit (92.5 minutes), a typical convergence time for magnetometer-only attitude estimation [45]. Because the EKF is a linearized algorithm, its estimates differ from those of the orthogonal filter at early times when the estimation errors are large. The EKF estimates agree quite well with those shown in Figure 2 at later times, however; the differences after 150 minutes being less than 0.003 deg in roll and yaw, 0.05 in pitch, 0.01 deg/hour in IRU drift bias, 1.5 nT in roll and yaw TAM bias, and 8 nT in pitch TAM bias. The smaller pitch estimation errors in Figure 1 compared to Figure 2 allow us to hope that the orthogonal filter would be more accurate than the EKF if its instability could be corrected.

CONCLUSION

The filter presented in this paper solves for the attitude matrix directly, along with an N-dimensional vector of bias states, by means of a new parameterization of the probability density function (PDF) of the combined attitude/bias state on the Cartesian product of SO(3) and $\mathbb{R}^N$. The negative-log-likelihood function, the negative of the logarithm of the PDF, is expressed as a quadratic function of the biases and, in parallel with the QUEST algorithm, a linear function of the attitude matrix. Maximum a posteriori probability estimates of the attitude matrix and the bias parameters are found by minimizing the negative-log-likelihood function. This filter can thus be seen as a generalization of the filter QUEST algorithm.

The filter uses Bayes's formula to incorporate measurement information, and it can accommodate more general measurement models than filter QUEST, in principle. The attitude matrix, which is equivalent to the lowest order $\ell = 1$ nontrivial irreducible representation of SO(3), is adequate for the simple measurement model considered in this paper, but more realistic measurement models may require the $\ell = 2$ irreducible representation to be included in the PDF, at the expense of increased computational complexity.

This filter uses the Fokker-Planck equation to propagate the PDF between measurements, so it incorporates process noise with more fidelity than does filter QUEST. The nonlinear Fokker-Planck equation introduces process-noise-dependent corrections to the implicit propagation of the estimates of the attitude and biases that are not considered by the extended Kalman filter.
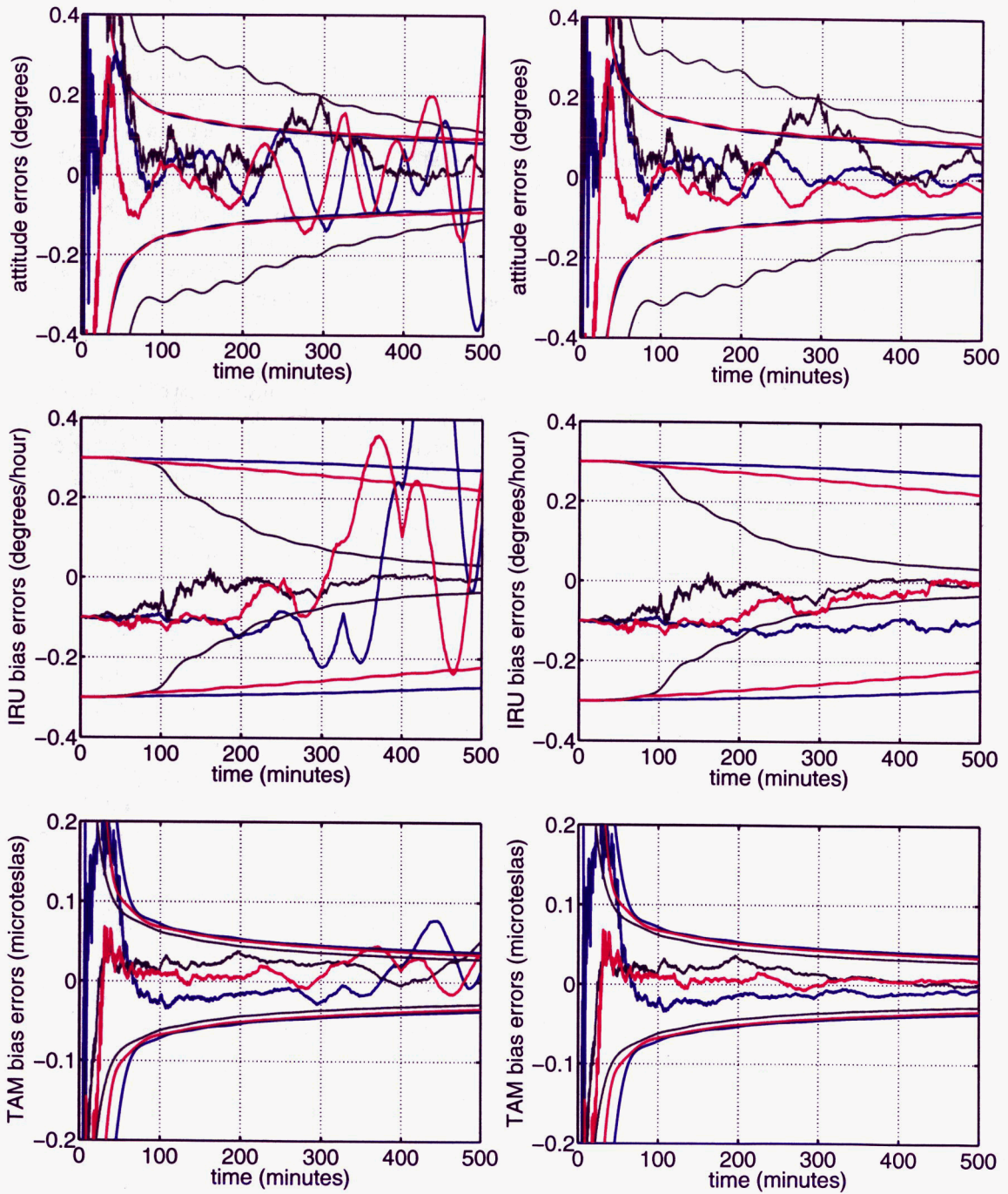
Figure 1. Orthogonal filter estimation errors and 3σ error bounds.
blue=roll, green=pitch, red=yaw

Figure 2. Orthogonal filter estimation errors and 3σ error bounds with $\Psi_b = 0$.
blue=roll, green=pitch, red=yaw

Because the Fokker-Planck equation is quadratic in the negative-log-likelihood function, it also introduces the higher-order $\ell = 2$ irreducible representation into the propagation. This order is not included in the PDF, so it must be ignored. Unfortunately, the resulting filter diverges without an *ad hoc* modification, and the modified filter performs no better than the extended Kalman filter. The neglect of the $\ell = 2$ terms is the most probable cause of the problem, so it is possible that adding $\ell = 2$ terms to the PDF would cure the instability. 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 can be found.

ACKNOWLEDGEMENTS

APPENDIX: THE ROTATION MATRIX $D^2(A)$

We would like to find a relationship similar to equation (9) for $D^2(A)$. Consider the real 6×6 matrix

$$R(A) \equiv \begin{bmatrix} R_{UL} & R_{UR} \\ R_{LL} & R_{LR} \end{bmatrix}, \tag{A1a}$$

with

$$R_{UL} \equiv \begin{bmatrix} A_{11}^2 & A_{12}^2 & A_{13}^2 \\ A_{21}^2 & A_{22}^2 & A_{23}^2 \\ A_{31}^2 & A_{32}^2 & A_{33}^2 \end{bmatrix}, \quad R_{LR} \equiv \begin{bmatrix} A_{22}A_{33} + A_{23}A_{32} & A_{23}A_{31} + A_{21}A_{33} & A_{21}A_{32} + A_{22}A_{31} \\ A_{32}A_{13} + A_{33}A_{12} & A_{33}A_{11} + A_{31}A_{13} & A_{31}A_{12} + A_{32}A_{11} \\ A_{12}A_{23} + A_{13}A_{22} & A_{13}A_{21} + A_{11}A_{23} & A_{11}A_{22} + A_{12}A_{21} \end{bmatrix}, \tag{A1bc}$$

$$R_{UR} \equiv \sqrt{2} \begin{bmatrix} A_{12}A_{13} & A_{13}A_{11} & A_{11}A_{12} \\ A_{22}A_{23} & A_{23}A_{21} & A_{21}A_{22} \\ A_{32}A_{33} & A_{33}A_{31} & A_{31}A_{32} \end{bmatrix}, \quad \text{and} \quad R_{LL} \equiv \sqrt{2} \begin{bmatrix} A_{21}A_{31} & A_{22}A_{32} & A_{23}A_{33} \\ A_{31}A_{11} & A_{32}A_{12} & A_{33}A_{13} \\ A_{11}A_{21} & A_{12}A_{22} & A_{13}A_{23} \end{bmatrix}, \tag{A1de}$$

where A is a member of SO(3). It is easy to see that

$$R(A^T) = R^T(A) \tag{A2}$$

and it can be shown by direct computation that $R(A)$ is a (reducible) representation of SO(3):

$$R(AA') = R(A)R(A'). \tag{A3}$$

These two properties taken together show that $R(A)$ is an orthogonal matrix, and it is not difficult to show that

$$R(\exp[-\boldsymbol{\theta}\times]) = \exp\begin{bmatrix} 0_{3\times3} & \sqrt{6}\,[\mathbf{u}_3\times]\,\text{diag}(\boldsymbol{\theta}) \\ \sqrt{6}\,\text{diag}(\boldsymbol{\theta})[\mathbf{u}_3\times] & [\boldsymbol{\theta}\times] \end{bmatrix}, \tag{A4}$$

where $\mathbf{u}_3$ is the unit vector

$$\mathbf{u}_3 \equiv (1/\sqrt{3})[1, 1, 1]^T . \tag{A5}$$

Substituting equation (9) into

$$D^2_{m'm}(A) = \sum_{\substack{m_1' m_2' \\ m_1 m_2}} C^{112}_{m_1' m_2' m'}\, C^{112}_{m_1 m_2 m}\, D^1_{m_1' m_1}(A)\, D^1_{m_2' m_2}(A) \tag{A6}$$

where $C^{j_1 j_2 j}_{m_1 m_2 m}$ denotes the Clebsch-Gordan (or Wigner) coefficient [27, 29], and exercising the symbolic manipulation capabilities of Mathematica® gives the desired relation between $D^2(A)$ and $R(A)$:

$$D^2(A) = M_2^\dagger R(A) M_2 , \tag{A7}$$

where

$$M_2 \equiv \begin{bmatrix} -1/2 & 0 & 1/\sqrt{6} & 0 & -1/2 \\ 1/2 & 0 & 1/\sqrt{6} & 0 & 1/2 \\ 0 & 0 & -2/\sqrt{6} & 0 & 0 \\ 0 & i/\sqrt{2} & 0 & i/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ -i/\sqrt{2} & 0 & 0 & 0 & i/\sqrt{2} \end{bmatrix} . \tag{A8}$$

The matrix M_2 obeys the relations

$$M_2^\dagger M_2 = I_{5\times5} \quad \text{and} \quad M_2 M_2^\dagger = I_{6\times6} - \mathbf{u}_6 \mathbf{u}_6^T , \tag{A9ab}$$

where

$$\mathbf{u}_6 \equiv \begin{bmatrix} \mathbf{u}_3 \\ \mathbf{0}_3 \end{bmatrix} \tag{A10}$$

is the basis for the one-dimensional null space of $M_2^\dagger$. The orthogonality of A can be used to show that $\mathbf{u}_6$ is an eigenvector of $R(A)$ with unity eigenvalue:

$$R(A)\mathbf{u}_6 = \mathbf{u}_6 , \tag{A11}$$

which leads to the relationship

$$R(A) = M_2 D^2(A) M_2^\dagger + \mathbf{u}_6 \mathbf{u}_6^T , \tag{A12}$$

showing that $R(A)$ contains $\ell = 0$ and $\ell = 2$ components, but no $\ell = 1$ component.

An $\ell = 2$ contribution to the real negative-log-likelihood function can be written in terms of a 5×5 array $\Psi_{\ell=2}$ of complex coefficients as

$$J_{A_\mu | Y \nu}^{\ell=2} = \tfrac{1}{2}\,\mathrm{tr}[\Psi_{\ell=2}^\dagger D^2(A)] = \tfrac{1}{2}\,\mathrm{tr}[\Psi_{\ell=2}^\dagger M_2^\dagger R(A) M_2] = \tfrac{1}{2}\,\mathrm{tr}[\Theta^T R(A)]\,, \tag{A13}$$

where

$$\Theta = M_2 \Psi_{\ell=2} M_2^\dagger \tag{A14}$$

is a 6×6 array of real coefficients. Because $N^\dagger \mathbf{u}_6 = \mathbf{0}_5$, this array obeys the constraints

$$\Theta \mathbf{u}_6 = \Theta^T \mathbf{u}_6 = \mathbf{0}_6\,, \tag{A15}$$

which mean that it has the 25 independent parameters appropriate to an $\ell = 2$ representation of SO(3). Violation of the constraints causes no problem; it merely mixes a superfluous $\ell = 0$ term into the negative-log-likelihood function.

Now we use these results to represent the term ignored by the approximation of equation (38), which is

$$\tfrac{1}{2}\mathbf{H}_\mu^T A^T F_\mu^m A \mathbf{H}_\mu - \tfrac{1}{6}\left|\mathbf{H}_\mu\right|^2 \mathrm{tr} F_\mu^m = \tfrac{1}{2}\,\mathrm{tr}\!\left(\mathcal{H} A^T \tilde{F}^m A\right), \tag{A16}$$

where $\mathcal{H}$ and $\tilde{F}^m$ are the symmetric traceless matrices

$$\mathcal{H} \equiv \mathbf{H}_\mu \mathbf{H}_\mu^T - \tfrac{1}{3}\left|\mathbf{H}_\mu\right|^2 I_{3\times 3} \quad \text{and} \quad \tilde{F}^m \equiv F_\mu^m - \tfrac{1}{3}(\mathrm{tr} F_\mu^m) I_{3\times 3}\,. \tag{A17}$$

Using the symmetry of $\mathcal{H}$ and $\tilde{F}^m$, equation (A16) can be represented in the form of equation (A13) with

$$\Theta \equiv \begin{bmatrix} \Theta_{UL} & \Theta_{UR} \\ \Theta_{LL} & \Theta_{LR} \end{bmatrix}, \tag{A18a}$$

where

$$\Theta_{UL} \equiv \begin{bmatrix} \mathcal{H}_{11}\tilde{F}_{11}^m & \mathcal{H}_{11}\tilde{F}_{22}^m & \mathcal{H}_{11}\tilde{F}_{33}^m \\ \mathcal{H}_{22}\tilde{F}_{11}^m & \mathcal{H}_{22}\tilde{F}_{22}^m & \mathcal{H}_{22}\tilde{F}_{33}^m \\ \mathcal{H}_{33}\tilde{F}_{11}^m & \mathcal{H}_{33}\tilde{F}_{22}^m & \mathcal{H}_{33}\tilde{F}_{33}^m \end{bmatrix}, \qquad \Theta_{LR} \equiv 2\begin{bmatrix} \mathcal{H}_{23}\tilde{F}_{23}^m & \mathcal{H}_{23}\tilde{F}_{31}^m & \mathcal{H}_{23}\tilde{F}_{12}^m \\ \mathcal{H}_{31}\tilde{F}_{23}^m & \mathcal{H}_{31}\tilde{F}_{31}^m & \mathcal{H}_{31}\tilde{F}_{12}^m \\ \mathcal{H}_{12}\tilde{F}_{23}^m & \mathcal{H}_{12}\tilde{F}_{31}^m & \mathcal{H}_{12}\tilde{F}_{12}^m \end{bmatrix}, \tag{A18bc}$$

$$\Theta_{UR} \equiv \sqrt{2}\begin{bmatrix} \mathcal{H}_{11}\tilde{F}_{23}^m & \mathcal{H}_{11}\tilde{F}_{31}^m & \mathcal{H}_{11}\tilde{F}_{12}^m \\ \mathcal{H}_{22}\tilde{F}_{23}^m & \mathcal{H}_{22}\tilde{F}_{31}^m & \mathcal{H}_{22}\tilde{F}_{12}^m \\ \mathcal{H}_{33}\tilde{F}_{23}^m & \mathcal{H}_{33}\tilde{F}_{31}^m & \mathcal{H}_{33}\tilde{F}_{12}^m \end{bmatrix}, \quad \text{and} \quad R_{LL} \equiv \sqrt{2}\begin{bmatrix} \mathcal{H}_{23}\tilde{F}_{11}^m & \mathcal{H}_{23}\tilde{F}_{22}^m & \mathcal{H}_{23}\tilde{F}_{33}^m \\ \mathcal{H}_{31}\tilde{F}_{11}^m & \mathcal{H}_{31}\tilde{F}_{22}^m & \mathcal{H}_{31}\tilde{F}_{33}^m \\ \mathcal{H}_{12}\tilde{F}_{11}^m & \mathcal{H}_{12}\tilde{F}_{22}^m & \mathcal{H}_{12}\tilde{F}_{33}^m \end{bmatrix}. \tag{A18de}$$

Note that Θ obeys equation (A15), because $\mathcal{H}$ and $\tilde{F}^m$ are traceless. The neglected terms in equations (87) and (89) can be similarly represented as $\ell = 2$ terms.

REFERENCES

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

2. Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970, Chapters 4–7

3. Schmidt, S. F., "The 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 Filtering," *Automatica*, Vol. 6, 1970, pp. 419–430

5. Murrell, J. W., "Precision Attitude Determination for Multimission Spacecraft," AIAA Paper AIAA-78-1248, *AIAA Guidance and Control Conference*, Palo Alto, CA, August 1978

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

7. Markley, F. L., "Attitude Error Representations for Kalman Filtering," *Journal of Guidance, Control, and Dynamics,* Vol. 26, No. 2, 2003, pp. 311–317

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

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

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

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

12. 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, 1995, pp. 1628-1632

13. Wan, E. A., and van der Merwe, R., "The Unscented Kalman Filter," Chapter 7 in *Kalman Filtering and Neural Networks*, ed. by S. S. Haykin, John Wiley & Sons, New York, 2001

14. Nørgaard, M, Poulsen, N. K., and Ravn, O., "New Developments in State Estimation for Nonlinear Systems," *Automatica*, Vol. 36, 2000, pp. 1627–1638

15. 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

16. Doucet, A., de Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, Springer, New York, 2001, Chapter 1

17. 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. 174–188

18. Crassidis, J. L., and Cheng, Y., "Particle Filtering for Sequential Spacecraft Attitude Estimation," AIAA Paper AIAA-2004-5337, *AIAA Guidance and Control Conference*, Providence, RI, August 2004

19. Oshman, Y., and Carmi, A., "Estimating Attitude from Vector Observations Using a Genetic Algorithm-Embedded Quaternion Particle Filter," AIAA Paper AIAA-2004-5340, *AIAA Guidance and Control Conference*, Providence, RI, August 2004

20. Gardiner, C. W., *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 3rd ed., Springer, Berlin, Heidelberg, New York, 2004, pp. 96–101

21. Daum, F. E., "New Exact Nonlinear Filters," Chapter 8 in *Bayesian Analysis of Time Series and Dynamical Models*, ed. By J. C. Spall, Marcel Dekker, New York, 1988

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

23. Lo, J. T.-H., and Eshelman, 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

24. Bauer, R., "Uniform Sampling of SO$_3$," *2001 Flight Mechanics Symposium*, NASA Conference Publication NASA/CP-2001-209986, Greenbelt, MD, 2001, pp. 347–359

25. Halmos, P. R., *Measure Theory*, D. van Nostrand Company, Inc., Princeton, 1950, pp. 250–263

26. Peter, F. and Weyl, H., "Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe," *Mathematische Annalen*, Vol. 97, 1927, pp. 737–755

27. Talman, J. D., *Special Functions: A Group Theoretical Approach, Based on Lectures by Eugene P. Wigner*, W. A. Benjamin, New York, 1968, Chapters 7 and 9

28. Gross, K. I., "On the Evolution of Noncommutative Harmonic Analysis," American Mathematical Monthly, Vol. 85, 1978, pp. 525–548

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

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

31. 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

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

33. Markley, F. L., and Mortari, D., "Quaternion Attitude Estimation Using Vector Observations," *Journal of the Astronautical Sciences*, Vol. 28, No. 2/3, 2000, pp. 359–380

34. Markley, F. L., "Attitude Determination Using Vector Observations and the Singular Value Decomposition," *Journal of the Astronautical Sciences*, Vol. 36, No. 3, 1988, pp. 245–258

35. Horn, R. A., and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, 1985, pp. 411–420, 18, 20

36. Markley, F. L., "Parameterization of the Attitude," Section 12.1 in *Spacecraft Attitude Determination and Control*, ed. By J. R. Wertz, D. Reidel, Dordrecht, Holland, 1978, p. 414

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

38. Markley, F. L., "Attitude Determination and Parameter Estimation Using Vector Observations: Theory," *Journal of the Astronautical Sciences,* Vol. 37, No. 1, 1989, pp. 41–58

39. Markley, F. L., "Attitude Determination and Parameter Estimation Using Vector Observations: Application," *Journal of the Astronautical Sciences,* Vol. 39, No. 3, 1991, pp. 367–381

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

41. Arnold, L., *Stochastic Differential Equations: Theory and Practice*, Wiley-Interscience, New York, 1974, pp. 163–175

42. 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

43. Andrews, S. F., and Bilanow, S., "Recent Flight Results of the TRMM Kalman Filter," AIAA Paper AIAA-2002-5047, *AIAA Guidance and Control Conference*, Monterey, CA, August 2002

44. Plett, M., "Magnetic Field Models," Appendix H in *Spacecraft Attitude Determination and Control*, ed. By J. R. Wertz, D. Reidel, Dordrecht, Holland, 1978, pp. 782, 783

45. Sedlak, J., and Hashmall, J., "Accurate Magnetometer/Gyroscope Attitudes Using a Filter with Correlated Sensor Noise," *Flight Mechanics Symposium 1997*, NASA Conference Publication 3345, Greenbelt, MD, 1976, pp. 83–87