The Journal of the Astronautical Sciences, Vol. 51, No. 1, January–March 2003, pp. 51–74

Constraint in Attitude Estimation Part I: Constrained Estimation[1]

Malcolm D. Shuster[2]

Felix qui potuit cognoscere de rerum causas.[3]

– Publius Vergilius Maro (70–19 B.C.E.)

Abstract

A complete and careful foundation is presented for maximum-likelihood attitude estimation and the calculation of measurement sensitivity matrices with the intent of revealing heretofore undisclosed pitfalls associated with unconstrained quaternion estimation. Efficient formulas are developed for computing the measurement sensitivity matrix for any attitude representation for which an efficient formula for the inverse kinematic equation is known. In particular, it is shown that the measurement sensitivity matrix for the quaternion is ambiguous and may take on a wide range of values. Hence, estimates of a quaternion which do not take correct account of the norm constraint will be physically meaningless. It is shown also that within Maximum Likelihood Estimation the form of the Wahba cost function for attitude estimation is incorrect when the attitude constraint is relaxed. A simple physical example is presented for quaternion estimation from noise-free vector measurements which fails when the norm constraint on the quaternion is relaxed. Part I of this work provides the basis for more detailed investigations of unconstrained attitude estimation in Part II [1].

Introduction

This is the first of two articles whose purpose is to dissuade practitioners of spacecraft attitude estimation from estimating a quaternion or quaternion correction which does not satisfy the appropriate norm constraint, at least to first order in the estimation error. This was also the intent of an earlier conference report of this work [2], where the intent was stated less bluntly.

[1]This and the succeeding article [1] are an expansion of an earlier conference report [2], presented in August 1993.
[2]Director of Research, Acme Spacecraft Company, 13017 Wisteria Drive, Box 328, Germantown, Maryland 20874. email: mdshuster@comcast.net.
[3]Translation: Happy [is he] who has been able to know the causes of things.

By "appropriate" we mean unit-norm for a quaternion which would be a *quaternion of rotation* if the norm constraint were enforced, and for a *quaternion correction* that the corrected quaternion have unit norm to this order. We say first order in the estimation error, because the linearization process of batch estimators or the extended Kalman filter discards terms of second order. In general, by *quaternion of rotation* [3] we mean an element of the quaternion group (i.e., homomorphic to $SO(3)$), whose domain is S^3, the unit sphere in four dimensions. By a *quaternion* in general, we mean an element of the quaternion skew field (algebra, division ring), whose domain is R^4. We will sometimes discard the tag "of rotation" when it is obvious from the context. The distinguishing feature of *quaternions*, of course, is the nature of the multiplication operation [3].

It has long been the feeling of this writer that unconstrained quaternion estimation[4] has been lacking in rigor and coherence and especially in analytical studies. The present work seeks to help fill this gap. The approach of this work is unsophisticated but careful. For the most part Part I provides the necessary background for Part II, but it contains original work as well. Our general investigatory tool is not the Kalman filter, which up to now has been the sole context of unconstrained quaternion estimation, but the covariance matrix or inverse covariance matrix (really, Fisher information matrix) of batch attitude estimation. We choose batch attitude estimation for the simple reason that it is more transparent than sequential estimation. This limits our studies to systems without process noise, hence, equivalently, to static systems, but the benefit of such studies will become very apparent in the sequel. In particular, the Kalman filter applied to a system without process noise must yield the same result within round-off error as a batch estimator. Hence, a failure of the batch estimator for such a system implies a failure of the Kalman filter for the same data.

We begin Part I with a careful presentation of maximum-likelihood attitude estimation. Next we present the attitude measurement sensitivity matrix in much more detail than it has been presented before. We present both constraint-sensitive and constraint-insensitive measurement sensitivity matrices since both have their place in correctly constrained attitude estimation. An important early result is that the value of the constraint-insensitive measurement sensitivity matrix for the quaternion is not unique, which has the immediate implication that the unconstrained quaternion estimate is also not unique and, therefore, without physical meaning. We next show that the Wahba problem is a maximum-likelihood attitude determination problem for direction measurements only for the correctly constrained estimation of the attitude. Finally, as a cliffhanger, we present a trivial but disturbing example of the failure of the unconstrained quaternion estimation process.

In Part II [1] of this paper we examine many examples of unconstrained attitude estimation and the restoration of the constraint. Again our attention is focused on batch attitude estimation, but we will also discuss the obvious consequences of our batch estimation results for the Kalman filter. An interesting outcome of our investigation is that there is genuine value in the estimation of an unconstrained attitude matrix, but none, it seems, in the estimation of an unconstrained quaternion. The reader should note that Part II assumes that the reader has read Part I.

[4]We postpone references to specific works until Part II.

Batch Attitude Estimation and Maximum Likelihood Estimation

The two principle ingredients of attitude estimation, apart from the data, are: (1) the attitude measurement model and (2) the estimation method. We assume generally that the attitude measurement can be written in the form

$$\mathsf{z} = \mathsf{f}(A) + \boldsymbol{\eta} \tag{1}$$

where z is the measurement vector, generally a column vector of dimension n. $\mathsf{f}(A)$ is a column-vector function of the attitude of the same dimension, which we may express generically as a function of the direction-cosine matrix [3]. $\boldsymbol{\eta}$ is the noise vector, which we generally assume to be Gaussian, zero-mean, and having an $n \times n$ covariance matrix R. We write $\boldsymbol{\eta} \sim \mathcal{N}(\mathsf{0}, R)$. It says something for the primitiveness of our field that we seldom (except for gyros) have need to go beyond such a simple measurement model.[5]

Given a set of N measurements, $\mathsf{z}_1, \ldots, \mathsf{z}_N$, which we assume depend only on the attitude (represented in our discussion by the attitude matrix), the *maximum-likelihood estimate* of the attitude is the value of A for which the probability density function of the measurements is a maximum. We write

$$A^*_{\text{ML}} = \arg \max_{A \varepsilon SO(3)} p(\mathsf{z}_1, \ldots, \mathsf{z}_n | A) \tag{2}$$

The probability density function (pdf) in this situation is generally called the *likelihood function* (hence, the name maximum-likelihood estimation) and written as $L(A \,|\, \mathsf{z}_1, \ldots, \mathsf{z}_N)$ to show the different emphasis. The vertical in the pdf in equation (2) may be interpreted either as showing just a functional dependence on A or a conditional pdf if A is a random matrix.

If the measurements are independent and Gaussian then the likelihood function can be written as

$$\begin{aligned} L(A) &= \prod_{k=1}^{N} \frac{1}{\sqrt{(2\pi)^{n_k} \det R_k}} \exp\left\{ -\frac{1}{2} [\mathsf{z}_k - \mathsf{f}_k(A)]^{\mathrm{T}} R_k^{-1} [\mathsf{z}_k - \mathsf{f}_k(A)] \right\} \\ &= \exp\left\{ -\frac{1}{2} \sum_{k=1}^{N} [[\mathsf{z}_k - \mathsf{f}_k(A)]^{\mathrm{T}} R_k^{-1} [\mathsf{z}_k - \mathsf{f}_k(A)] + \log \det R_k + n_k \log 2\pi] \right\} \end{aligned} \tag{3}$$

where we have not written the dependence of $L(A)$ on the measurements explicitly. If we now define the *negative-log-likelihood function* $J_{\text{NLL}}(A)$ as (for convenience we no longer write the measurements as arguments)

$$\begin{aligned} J_{\text{NLL}}(A) &= -\log L(A) \\ &= \frac{1}{2} \sum_{k=1}^{N} \{[\mathsf{z}_k - \mathsf{f}_k(A)]^{\mathrm{T}} R_k^{-1} [\mathsf{z}_k - \mathsf{f}_k(A)] + \log \det R_k + n_k \log 2\pi\} \end{aligned} \tag{4}$$

We may write equivalently that

$$A^*_{\text{ML}} = \arg \min_{A \varepsilon SO(3)} J_{\text{NLL}}(A) \tag{5}$$

[5]Generally, but not without exception, column arrays will be denoted by boldface sans serif type, matrices by upper-case italic Roman type, scalars (other than elements of arrays) by lower-case italic Roman type. Quaternions will always be identified by an overbar. By *Roman* we mean the type style, not the alphabet.

that is, the maximum-likelihood estimate minimizes the negative-log-likelihood function. We note finally, that if we are not estimating parameters which characterize the measurement-model covariances, then the last two terms do not affect the minimization, and we have finally that the maximum-likelihood estimate of the attitude must minimize the least-squares cost function

$$J(A) = \frac{1}{2}\sum_{k=1}^{N}[\mathbf{z}_k - \mathbf{f}_k(A)]^{\mathrm{T}}R_k^{-1}[\mathbf{z}_k - \mathbf{f}_k(A)] \tag{6}$$

which, in MLE parlance, is the *data-dependent part of the negative logarithm of the likelihood function* [4]. Within MLE, there is no arbitrary constant factor nor arbitrary constant term in the cost function. The cost function of equation (6) will be the basis of our estimation program.[6]

Since the attitude matrix A, or any other attitude representation, depends on only three independent parameters, it follows that $J(A)$ is sensitive to only three independent attitude parameters. If, following ancient practice, we write $A = A(\boldsymbol{\phi})$, with $\boldsymbol{\phi}$, say, a 3×1 matrix of the Euler angles [3], we could then minimize a batch weighted least-squares cost function

$$J(\boldsymbol{\phi}) \equiv J(A(\boldsymbol{\phi})) = \frac{1}{2}\sum_{k=1}^{N}[\mathbf{z}_f - \mathbf{f}_k(A(\boldsymbol{\phi}))]^{\mathrm{T}}R_k^{-1}[\mathbf{z}_f - \mathbf{f}_k(A(\boldsymbol{\phi}))] \tag{7}$$

The minimizing value of $\boldsymbol{\phi}$, the estimate $\boldsymbol{\phi}^*$, is usually found by the Newton-Raphson method [4], which follows.

If the desired estimate does not lie on the boundary of the domain of $\boldsymbol{\phi}$ (usually because the domain has no boundary), then[7]

$$\frac{\partial J}{\partial \boldsymbol{\phi}^{\mathrm{T}}}(\boldsymbol{\phi}^*) = \mathbf{0} \tag{8}$$

which must be solved iteratively.

If $\boldsymbol{\phi}_i^*(-)$ is an approximate (*a priori*) value of $\boldsymbol{\phi}^*$ at step i, then in some small region containing both $\boldsymbol{\phi}_i^*(-)$ and $\boldsymbol{\phi}^*$ we have (writing $\boldsymbol{\phi} = \boldsymbol{\phi}_i^*(-) + \Delta\boldsymbol{\phi}_i$)

$$\frac{\partial J}{\partial \boldsymbol{\phi}^{\mathrm{T}}}(\boldsymbol{\phi}) = \frac{\partial J}{\partial \boldsymbol{\phi}^{\mathrm{T}}}(\boldsymbol{\phi}_i^*(-) + \Delta\boldsymbol{\phi}_i) = \frac{\partial J}{\partial \boldsymbol{\phi}^{\mathrm{T}}}(\boldsymbol{\phi}_i^*(-)) + \frac{\partial^2 J}{\partial \boldsymbol{\phi}^{\mathrm{T}}\partial \boldsymbol{\phi}}(\boldsymbol{\phi}_i^*(-))\Delta\boldsymbol{\phi}_i + \ldots \tag{9}$$

Truncating the Taylor series at linear order in $\Delta\boldsymbol{\phi}_i$, setting the gradient of $J(\boldsymbol{\phi}_i^*(-) + \Delta\boldsymbol{\phi}_i)$ equal to $\mathbf{0}$ at $\Delta\boldsymbol{\phi}_i^*(+)$, and solving for $\Delta\boldsymbol{\phi}_i^*(+)$ lead to

$$\Delta\boldsymbol{\phi}_i^*(+) = -\left[\frac{\partial^2 J}{\partial \boldsymbol{\phi}^{\mathrm{T}}\partial \boldsymbol{\phi}}(\boldsymbol{\phi}_i^*(-))\right]^{-1}\frac{\partial J}{\partial \boldsymbol{\phi}^{\mathrm{T}}}(\boldsymbol{\phi}_i^*(-)) \tag{10}$$

[6]There is a minor complication in the form of the likelihood function because it will turn out the $\boldsymbol{\eta}_k$ can have only a finite range while a Gaussian random variable ranges over all space. This consideration is treated in detail in reference [5] and can be ignored here, essentially because the Gaussian distribution falls off so quickly.

[7]We use the convention that the derivative of a scalar function with respect to a column vector is a row vector, hence, the differentiation with respect to a row vector in equation (8). The nabla operator, however, acting on a scalar function always produces a column vector. Note that local minima will also satisfy equation (8).

and the $(i + 1)$-th *a priori* value for the next iteration is the *a posteriori* value at step i

$$\boldsymbol{\phi}^*_{i+1}(-) = \boldsymbol{\phi}^*_i(+) = \boldsymbol{\phi}^*_i(-) + \Delta\boldsymbol{\phi}^*_i(+) \tag{11}$$

(Note in passing that necessarily $\Delta\boldsymbol{\phi}^*_i(-) = \mathbf{0}$.) If one is sufficiently close to the estimate, then by continuing this process *ad infinitum*

$$\lim_{i\to\infty} \boldsymbol{\phi}^*_i(+) = \boldsymbol{\phi}^* \tag{12}$$

and in the limit that the quantity of data becomes infinite (the asymptotic limit)

$$\lim_{N\to\infty} \boldsymbol{\phi}^* = \boldsymbol{\phi}^{\text{true}} \tag{13}$$

Generally, a different method is required to supply the initial estimate of $\boldsymbol{\phi}$.

The covariance matrix of the estimate in the asymptotic limit is given by

$$P_{\boldsymbol{\phi\phi}} \equiv E\{(\boldsymbol{\phi}^* - \boldsymbol{\phi}^{\text{true}})(\boldsymbol{\phi}^* - \boldsymbol{\phi}^{\text{true}})^{\mathrm{T}} = F^{-1}_{\boldsymbol{\phi\phi}} \tag{14}$$

where $F_{\boldsymbol{\phi\phi}}$, the *Fisher information matrix*, is given by

$$F_{\boldsymbol{\phi\phi}} \equiv E\left\{\frac{\partial^2 J}{\partial\boldsymbol{\phi}^{\mathrm{T}}\partial\boldsymbol{\phi}}(\boldsymbol{\phi}^{\text{true}})\right\} = E\left\{\frac{\partial J}{\partial\boldsymbol{\phi}^{\mathrm{T}}}(\boldsymbol{\phi}^{\text{true}})\frac{\partial J}{\partial\boldsymbol{\phi}}(\boldsymbol{\phi}^{\text{true}})\right\} \tag{15}$$

which, for practical reasons, is generally evaluated at $\boldsymbol{\phi}^*$. Unless the Fisher information matrix is full-rank, the covariance matrix will not exist (nor will ϕ^*). Nonetheless, we will frequently write $P^{-1}_{\boldsymbol{\phi\phi}}$ for the Fisher information matrix, even when $P_{\boldsymbol{\phi\phi}}$ does not exist.

A variant of the Newton-Raphson method is the *Gauss-Newton Method* [4]. In this method the Hessian matrix in equation (10) is replaced by its expectation, the Fisher information matrix. The Fisher information matrix is generally a much smoother function of the parameters than the Hessian matrix and the estimation sequence generally converges faster.

We may write the measurement as

$$\begin{aligned} \mathbf{z}_k &= \mathbf{f}_k(A(\boldsymbol{\phi}^*_i(-))) + \frac{\partial\mathbf{f}_k}{\partial\boldsymbol{\phi}^{\mathrm{T}}}(A(\boldsymbol{\phi}^*_i(-)))\,\Delta\boldsymbol{\phi}_i + \ldots + \boldsymbol{\eta}_k \\ &= \mathbf{z}_{k,i}(A(\boldsymbol{\phi}^*_i(-))) + H_{k,i}(A(\boldsymbol{\phi}^*_i(-)))\,\Delta\boldsymbol{\phi}_i + \ldots + \boldsymbol{\eta}_k \end{aligned} \tag{16}$$

$\mathbf{z}_{k,i}$ is the *a priori* value of the measurement (given $\boldsymbol{\phi}^*_i(-)$) and $H_{k,i}$ is the *measurement sensitivity matrix*. $\Delta\mathbf{z}_{k,i} \equiv \mathbf{z}_k - \mathbf{z}_{k,i}$ is called the *residual measurement*. In this work, k will always denote the measurement index and i the iteration index. Sometimes we will write H_ϕ or $H_{\boldsymbol{\phi},k,i}$ to emphasize the attitude representation. In terms of these quantities

$$P^{-1}_{\boldsymbol{\phi\phi},i} = \sum_{k=1}^{N} H^{\mathrm{T}}_{k,i} R^{-1}_k H_{k,i}, \qquad \mathbf{i}^*_i(+) = \sum_{k=1}^{N} H^{\mathrm{T}}_{k,i} R^{-1}_k \Delta\mathbf{z}_{k,i} \tag{17ab}$$

$$\Delta\boldsymbol{\phi}^*_i(+) = P_{\boldsymbol{\phi\phi},i}\,\mathbf{i}^*_i(+), \quad \boldsymbol{\phi}^*_i(+) = \boldsymbol{\phi}^*_i(-) + \Delta\boldsymbol{\phi}^*_i(+) = \boldsymbol{\phi}^*_{i+1}(-) \tag{17cd}$$

and

$$P_{\boldsymbol{\phi\phi}} = \lim_{i\to\infty} P_{\boldsymbol{\phi\phi},i}, \qquad \boldsymbol{\phi}^* = \lim_{i\to\infty} \boldsymbol{\phi}^*_i(+) \tag{18ab}$$

The vector $\mathbf{i}_i^*(+)$ is generally called the *information vector*.

Equations (16) through (18) were standard practice twenty-five years ago [6]. These three-dimensional parameterizations, however, are not regular everywhere and are inconvenient for practical calculations because of the complicated functions which must be differentiated. In addition, there is usually a choice for the space coordinate axes for which the attitude error covariance matrix will become infinite, even if the attitude is known with great precision. These difficulties are eliminated by the methods of the next section.

The Attitude Increment Vector

Consider now a parameterization of different character, which some workers choose to call "multiplicative," whereas the former choice was "additive." Given the *a priori* estimate of the attitude matrix $A_i^*(-)$ we write now

$$A = (\delta A_i)A_i^*(-) = \exp\{[\![\boldsymbol{\epsilon}_i]\!]\}A_i^*(-) \tag{19}$$

where $\boldsymbol{\epsilon}_i$, *the attitude increment vector*, is the rotation vector of a very small rotation δA_i. Here, $[\![\mathbf{v}]\!]$ is the 3×3 antisymmetric matrix [3]

$$[\![\mathbf{v}]\!] \equiv \begin{bmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & v_1 \\ v_2 & -v_1 & 0 \end{bmatrix} \tag{20}$$

and $\exp\{\cdot\}$ is the matrix exponential function. Euler's formula in terms of $\boldsymbol{\epsilon}$ becomes

$$\delta A(\boldsymbol{\epsilon}) = \exp\{[\![\boldsymbol{\epsilon}]\!]\} = \cos|\boldsymbol{\epsilon}|\, I_{3\times3} + \frac{1 - \cos|\boldsymbol{\epsilon}|}{|\boldsymbol{\epsilon}|^2}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}} + \frac{\sin|\boldsymbol{\epsilon}|}{|\boldsymbol{\epsilon}|}[\![\boldsymbol{\epsilon}]\!] \tag{21}$$

Because $\boldsymbol{\epsilon}$ is very small, we may write equation (21) as

$$\delta A = I_{3\times3} + [\![\boldsymbol{\epsilon}]\!] + O(|\boldsymbol{\epsilon}|^2) \tag{22}$$

where the matrix $I_{3\times3}$ is the 3×3 identity matrix. Obviously, $\delta A_i(-)$ has the value $I_{3\times3}$ and $\boldsymbol{\epsilon}_i(-)$ the value $\mathbf{0}$.

We now define the sensitivity matrix in terms of $\boldsymbol{\epsilon}$

$$H_{\epsilon,i} \equiv \frac{\partial \mathbf{f}((\delta A(\boldsymbol{\epsilon}))A_i^*(-))}{\partial \boldsymbol{\epsilon}^{\mathrm{T}}}(\boldsymbol{\epsilon}_i) \tag{23}$$

and proceed with the Newton-Raphson or Gauss-Newton method in the usual way. The covariance matrix is defined as

$$P_{\epsilon\epsilon} \equiv E\{\tilde{\boldsymbol{\epsilon}}^*\tilde{\boldsymbol{\epsilon}}^{*\mathrm{T}}\} \tag{24}$$

with $\tilde{\boldsymbol{\epsilon}}^*$, the *attitude error vector*, defined by[8]

$$A^* \equiv e^{[\![\tilde{\boldsymbol{\epsilon}}^*]\!]}A^{\mathrm{true}} \tag{25}$$

The attitude error covariance matrix defined in terms of $\tilde{\boldsymbol{\epsilon}}^*$ has the advantage that it is independent of the choice of space axes, and if the body axes are rotated by a rotation with direction-cosine matrix C, then simply

$$\tilde{\boldsymbol{\epsilon}} \to \tilde{\boldsymbol{\epsilon}}' = C\tilde{\boldsymbol{\epsilon}} \quad \text{and} \quad P_{\epsilon\epsilon} \to P'_{\epsilon\epsilon} = CP_{\epsilon\epsilon}C^{\mathrm{T}} \tag{26ab}$$

[8]We use the tilde to avoid confusing the attitude error vector with the estimate of the attitude increment.

Clearly, there is no singularity of $P_{\epsilon\epsilon}$ associated with the choice of coordinate axes. The "mechanics" of the attitude error and attitude increment vectors are identical.

The Measurement Sensitivity Matrix for the Attitude Increment Vector

Equation (1) for the attitude measurement model is too general for practical use. In almost every practical case an attitude measurement consists of a function of inner products of the components of a direction in space with respect to a coordinate system fixed in the spacecraft body. Consider, for example, an arbitrary physical vector $\boldsymbol{v}$ with 3×1 matrix representations $\mathbf{v}_{\mathcal{B}}$ and $\mathbf{v}_{\mathcal{I}}$ with respect to the body and inertial coordinate systems, respectively. If A, the attitude matrix, transforms representations from the inertial to the body coordinate system, then

$$\mathbf{v}_{\mathcal{B}} = A\mathbf{v}_{\mathcal{I}} \tag{27}$$

For a three-axis magnetometer, the measurement consist ideally of the three components of the magnetic field in the magnetometer frame. Hence, if $\boldsymbol{v}$ is the magnetic field vector, we may write the vector measurement as[9]

$$\mathbf{z} = \begin{bmatrix} \hat{x}\cdot\boldsymbol{v} \\ \hat{y}\cdot\boldsymbol{v} \\ \hat{z}\cdot\boldsymbol{v} \end{bmatrix} + \boldsymbol{\eta} = \begin{bmatrix} \hat{\mathbf{x}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}} \\ \hat{\mathbf{y}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}} \\ \hat{\mathbf{z}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}} \end{bmatrix} + \boldsymbol{\eta} \tag{28}$$

where $\hat{x}$, $\hat{y}$ and $\hat{z}$ are the three coordinate axes of the magnetometer, with body representations $\hat{\mathbf{x}}_B$, $\hat{\mathbf{y}}_B$ and $\hat{\mathbf{z}}_B$, and $\boldsymbol{\eta}$ is the measurement noise vector, generally assumed to be Gaussian. For a focal-plane sensor, typified by vector Sun sensors and star trackers, the measurement takes the form of two scalar measurements

$$\zeta_1 = \frac{\hat{\mathbf{x}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}}}{\hat{\mathbf{z}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}}} + \eta_1, \quad \text{and} \quad \zeta_2 = \frac{\hat{\mathbf{y}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}}}{\hat{\mathbf{z}}_B^{\mathrm{T}} A\mathbf{v}_{\mathcal{I}}} + \eta_2, \tag{29ab}$$

where now $\boldsymbol{v}$ is the unit vector of the sensed direction. The common component of all of these measurements are scalars of the form $\mathbf{u}^{\mathrm{T}}A\mathbf{v}$. We take, therefore, as our basic scalar measurement

$$z = \mathbf{u}^{\mathrm{T}}A\mathbf{v} + \eta \tag{30}$$

Substituting equations (19) and (22) into equation (30) leads to[10]

$$\begin{aligned} z &\approx \mathbf{u}^{\mathrm{T}}A(-)\mathbf{v} + \mathbf{u}^{\mathrm{T}}[\![\boldsymbol{\epsilon}]\!]A(-)\mathbf{v} + \eta \\ &= \mathbf{u}^{\mathrm{T}}A(-)\mathbf{v} + (\mathbf{u}\times A(-)\mathbf{v})^{\mathrm{T}}\boldsymbol{\epsilon} + \eta \\ &\equiv z_o(A(-)) + H_\epsilon(A(-))\boldsymbol{\epsilon} + \eta \end{aligned} \tag{31}$$

and we have used the fact that

$$[\![\boldsymbol{\epsilon}]\!]A(-)\mathbf{v} = -[\![A(-)\mathbf{v}]\!]\boldsymbol{\epsilon} \tag{32}$$

[9]Equation (28) is equivalent to $\mathbf{z} = S^{\mathrm{T}}A\mathbf{v}_I + \boldsymbol{\eta} \equiv \mathbf{z}_{\text{sensor}}$, where S is the alignment matrix [7]. Generally, however, when we write $\mathbf{z}$ we will mean $\mathbf{z}_B = S\mathbf{z}_{\text{sensor}}$, the measurement vector in body coordinates, when dealing with vector measurements, as in equation (37) below.

[10]We will frequently suppress subscripts and superscripts to make our notation less cumbersome. This will often include the asterisk on estimates, when the idea is already conveyed by $(\pm)$.

Thus, the sensitivity matrix $H_\epsilon(A(-))$ for a simple scalar measurement is given by the 1×3 matrix

$$^{(\mathrm{scalar})}H_\epsilon(A(-)) = (\mathbf{u} \times A(-)\mathbf{v})^{\mathrm{T}} \tag{33}$$

The *a posteriori* value $\boldsymbol{\epsilon}_i(+)$ is that obtained by minimizing the batch least-squares cost function over $\boldsymbol{\epsilon}_i$, after which the *a posteriori* value of the attitude[11] at the i-th iteration is $A_i(+) = \delta A_i(+)A_i(-) = \delta A(\boldsymbol{\epsilon}_i(+))A_i(-) = A_{i+1}(-)$. This approach must still contend with the presence of local minima but no longer suffers from possible singular points of the Euler angles and leads to a much simpler measurement sensitivity matrix than that for the Euler angles, because it avoids the need to differentiate the attitude matrix.

If the measurement is an arbitrary scalar function of the representation of the measured vector in body coordinates, then we may write successively

$$\begin{aligned} z &= f(A\mathbf{v}) + \eta = f(A(-)\mathbf{v} - [\![A(-)\mathbf{v}]\!]\boldsymbol{\epsilon}) + \eta \\ &= f(A(-)\mathbf{v}) + [(\nabla f(A(-)\mathbf{v})) \times (A(-)\mathbf{v})]^{\mathrm{T}}\boldsymbol{\epsilon} + \eta \end{aligned} \tag{34}$$

which amounts to replacing $\mathbf{u}$ by $\nabla f(A(-)\mathbf{v})$ in the earlier equations.[12] This allows us to accommodate focal-plane measurements as given by equations (29) within our model. Thus, for example, if

$$f(A\mathbf{v}) = \frac{\mathbf{a}^{\mathrm{T}}A\mathbf{v}}{\mathbf{b}^{\mathrm{T}}A\mathbf{v}} \tag{35}$$

as in equations (29), then the effective value of $\mathbf{u}$ is given by

$$\mathbf{u} = \nabla f(A(-)\mathbf{v}) = (\mathbf{b}^{\mathrm{T}}A(-)\mathbf{v})^{-2}(\mathbf{b} \times \mathbf{a}) \times (A(-)\mathbf{v}) \tag{36}$$

For the special case that $\mathbf{z}$ is a 3×1 matrix of the three components of a vector $\mathbf{w}$ *in the body frame*, $\mathbf{w} = A\mathbf{v}$, then simply

$$\mathbf{z} = A\mathbf{v} + \boldsymbol{\eta} = A(-)\mathbf{v} - [\![A(-)\mathbf{v}]\!]\boldsymbol{\epsilon} + \boldsymbol{\eta} \tag{37}$$

a form which has been used to advantage in other studies [8, 9]. The measurement sensitivity matrix for a vector measurement is then

$$^{(\mathrm{vector})}H_\epsilon(A(-)) = -[\![A(-)\mathbf{v}]\!] \tag{38}$$

The sensitivity matrices of equations (33) and (38) are considerably simpler than the corresponding sensitivity matrices obtained from differentiating the measurement function in terms of the Euler angles for a finite rotation.

One constructs the sensitivity matrix for a two-dimensional measurement from focal-plane sensors in a similar manner making use of equation (36). (Note that η_1 and η_2 need not be statistically independent.) However, in practice it is easier to convert these focal-plane measurements into unit vectors [8]. (See also Appendix A.)

While it is clear from equation (19) that $\boldsymbol{\epsilon}$ must be the rotation vector, when $\boldsymbol{\epsilon}$ is small we prefer to refer to it simply as the attitude increment vector. The reason for this is that many three-parameter representations of the attitude differ from the infinitesimal rotation vector only by terms of $O(|\boldsymbol{\epsilon}|^2)$ when $\boldsymbol{\epsilon}$ is very small. Hence, the

[11]While only the constant and linear terms in $\boldsymbol{\epsilon}$ are used in the linear approximation of the measurement, the higher-order terms cannot be neglected in computing δA from $\boldsymbol{\epsilon}$.
[12]Unless otherwise noted, the gradient will always be with respect to the entire argument of the function.

measurement sensitivity matrix for these other representations[13] when $\boldsymbol{\epsilon} \approx \mathbf{0}$ will be the same. These other representations include any asymmetric set of Euler angles, twice the Rodrigues vector, four times the modified Rodrigues vector, and also twice the vector components of the quaternion [3]. Thus, it seems pointless to specify which of the possible infinitesimal attitude representations is intended by $\boldsymbol{\epsilon}$. In updating the attitude one must specify the exact nature of $\boldsymbol{\epsilon}$ so that a finite parameterization of the spacecraft attitude can be chosen. The choice, in general, is inconsequential except for the computational burden.[14]

Given this freedom of choice in the definition of $\boldsymbol{\epsilon}$, let us define it to be the array of 1-2-3 Euler angles, measured from the most recently estimated body axes, while $\boldsymbol{\phi}$ is the same as before, but we may specify it too as a 1-2-3 sequence. Then we obtain the same equations for attitude estimation as before, but the choice between $\boldsymbol{\phi}$ and $\boldsymbol{\epsilon}$ becomes simply one of frame. We see this also in the measurement sensitivity matrices where an inertial representation $\mathbf{V}$ is always proceeded by the *a priori* estimate of the attitude $A_i(-)$, transforming it to a representation with respect to the predicted body axes. The same situation would have occurred had we chosen the representation to be the rotation vector or twice the Rodrigues vector or four times the modified Rodrigues vector [3]. The only appreciable difference between inertial updates or predicted-body-referenced updates is the computational burden, which has been shown to be less in the latter, or problems with singularities, which are greater in the former. We shall see a similar trade-off in the Kalman filter in Part II.

Alternate Update Strategies

The quaternion, likewise, can be expressed in terms of $\boldsymbol{\epsilon}$ as

$$\bar{q} = \delta\bar{q}(\boldsymbol{\epsilon}) \otimes \bar{q}(-) = \begin{bmatrix} \dfrac{\sin(|\boldsymbol{\epsilon}|/2)}{|\boldsymbol{\epsilon}|/2}\dfrac{\boldsymbol{\epsilon}}{2} \\ \cos(|\boldsymbol{\epsilon}|/2) \end{bmatrix} \otimes \bar{q}(-) \approx \begin{bmatrix} \boldsymbol{\epsilon}/2 \\ 1 \end{bmatrix} \otimes \bar{q}(-) \tag{39}$$

with the attitude matrix now a function of the quaternion [3]. The update is still carried out in terms of $\boldsymbol{\epsilon}$, but the attitude is accumulated as the quaternion. This has the advantage of avoiding the six constraints for the attitude matrix in favor of only one for the quaternion, while avoiding the singularities of the other attitude representations. The inevitable violation of the quaternion norm constraint due to round-off error is easily repaired for the quaternion by division by the norm [11]. Restoring orthogonality to a corrupted direction-cosine matrix is far more complicated [12–16]. (See also Part II.)

The use of $\boldsymbol{\epsilon}$ in calculating the attitude correction is not the only possible approach. We may also write the correction of the direction-cosine matrix in the form

$$A = A(-) + \Delta A \tag{40}$$

and the quaternion correction in the form

$$\bar{q} = \bar{q}(-) + \Delta\bar{q} \tag{41}$$

[13]A sensitivity matrix *for a particular attitude representation* focuses on the parameters to which the measurement is sensitive. It is, however, always a measurement sensitivity matrix.

[14]Markley [10] has pointed out that except for the Euler angles even the second-order 3×3 scalar measurement sensitivity matrices for any infinitesimal three-dimensional representation of the attitude are the same as that for $\boldsymbol{\epsilon}$.

Estimation strategies using the additive corrections for attitude representations of dimension higher than three suffer from several drawbacks. Firstly, the four elements of the quaternion and, even more so, the nine elements of the direction-cosine matrix, are constrained. Hence, ΔA and $\Delta\bar{q}$ must be constrained values for the left member, at least to first order. There is no mathematical justification for relaxing the constraint on the attitude estimate.

The Quaternion Measurement Sensitivity Matrix

Having warned against carrying out estimation procedures in terms of the four components of the quaternion, we now develop measurement sensitivity matrices for such a procedure, but also, as we shall see, for correctly constrained estimation. To develop a measurement sensitivity matrix in terms of the additive quaternion correction we write (for a simple scalar measurement)

$$z = \mathbf{u}^{\mathrm{T}} A(\bar{q})\mathbf{v} + \eta \tag{42}$$

with now

$$A(\bar{q}) = (q_4^2 - |\mathbf{q}|^2) I_{3\times 3} + 2\mathbf{q}\mathbf{q}^{\mathrm{T}} + 2q_4[\![\mathbf{q}]\!] \tag{43}$$

and

$$\bar{q} = \begin{bmatrix} \mathbf{q} \\ q_4 \end{bmatrix} \tag{44}$$

Then

$$H_{\bar{q}}^{*}(\bar{q}(-)) \equiv \left.\frac{\partial z}{\partial \bar{q}^{\mathrm{T}}}\right|_{\bar{q}(-)} = \mathbf{u}^{\mathrm{T}} \left.\frac{\partial A(\bar{q})}{\partial \bar{q}^{\mathrm{T}}}\right|_{\bar{q}(-)} \mathbf{v} \tag{45}$$

Unlike the three-column sensitivity matrices of the previous section, these sensitivity matrices are insensitive to the quaternion norm constraint. This is because equation (45) treats all four components of the quaternion as independent variables. To emphasize that difference, such sensitivity matrices are distinguished by an asterisk.

Explicit differentiation leads to

$$\begin{aligned} H_{\bar{q}}^{*}(\bar{q})\,\Delta\bar{q} = 2[&(\mathbf{u}\cdot\mathbf{v})(q_4\Delta q_4 - \mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}) + \mathbf{q}^{\mathrm{T}}(\mathbf{u}\mathbf{v}^{\mathrm{T}} + \mathbf{v}\mathbf{u}^{\mathrm{T}})\,\Delta\mathbf{q} \\ &+ (\mathbf{u}\times\mathbf{v})\cdot\mathbf{q}\,\Delta q_4 + q_4(\mathbf{u}\times\mathbf{v})^{\mathrm{T}}\Delta\mathbf{q}] \end{aligned} \tag{46}$$

where for convenience we have discarded the designation $(-)$ of $\bar{q}(-)$ for the moment. If we write now

$$H_{\bar{q}}^{*}(\bar{q}) = [\mathbf{h}^{*\mathrm{T}}(\bar{q})\,|\,h_4^{*}(\bar{q})] \equiv \bar{\mathbf{h}}^{*\mathrm{T}}(\bar{q}) \tag{47}$$

then

$$\bar{\mathbf{h}}^{*}(\bar{q}) = 2\begin{bmatrix} -(\mathbf{u}\cdot\mathbf{v})\mathbf{q} + (\mathbf{u}\mathbf{v}^{\mathrm{T}} + \mathbf{v}\mathbf{u}^{\mathrm{T}})\mathbf{q} + q_4(\mathbf{u}\times\mathbf{v}) \\ (\mathbf{u}\cdot\mathbf{v})q_4 + (\mathbf{u}\times\mathbf{v})\cdot\mathbf{q} \end{bmatrix} = -2M(\mathbf{u},\mathbf{v})\bar{q} \tag{48}$$

The matrix M is symmetric and traceless and can be factored as

$$M(\mathbf{u},\mathbf{v}) = \begin{bmatrix} [\![\mathbf{u}]\!] & \mathbf{u} \\ -\mathbf{u}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} -[\![\mathbf{v}]\!] & \mathbf{v} \\ -\mathbf{v}^{\mathrm{T}} & 0 \end{bmatrix} \tag{49}$$

If we define the quaternion representations of the three-vectors by

$$\bar{\mathbf{u}} \equiv \begin{bmatrix} \mathbf{u} \\ 0 \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{v}} \equiv \begin{bmatrix} \mathbf{v} \\ 0 \end{bmatrix} \tag{50}$$

then we can write

$$M(\mathbf{u},\mathbf{v}) = \{\bar{\mathbf{u}}\}_L \{\bar{\mathbf{v}}\}_R \tag{51}$$

where $\{\cdot\}_L$ and $\{\cdot\}_R$ are the matrix representations of the quaternion multiplication rules [3]

$$\bar{p} \otimes \bar{q} = \{\bar{p}\}_L \bar{q} = \{\bar{q}\}_R \bar{p} \tag{52}$$

Combining equations (49) through (52) leads finally to

$$^{(\text{scalar})}H_{\bar{q}}^{*}(\bar{q}(-)) = -2[\bar{\mathbf{u}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}}]^{\mathrm{T}} \tag{53}$$

The result can be extended to more general measurement models by replacing $\mathbf{u}$ by $\nabla f(A(-)\mathbf{v})$ as in equation (34).

For the case of a complete vector measurement, we have

$$\mathbf{z} = A(\bar{q})\mathbf{v} + \boldsymbol{\eta} \tag{54}$$

We may regard this as a 3×1 matrix of measurements of the form given by equation (30). Thus, applying equation (53) to the three components of $\mathbf{z}$ leads to

$$\begin{aligned} ^{(\text{vector})}H_{\bar{q}}^{*\mathrm{T}}(\bar{q}(-)) &= -2 \begin{bmatrix} (\bar{\mathbf{1}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}})^{\mathrm{T}} \\ (\bar{\mathbf{2}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}})^{\mathrm{T}} \\ (\bar{\mathbf{3}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}})^{\mathrm{T}} \end{bmatrix} \\ &= -2[\bar{\mathbf{1}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}} | \bar{\mathbf{2}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}} | \bar{\mathbf{3}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}}]^{\mathrm{T}} \end{aligned} \tag{55}$$

where

$$\hat{\mathbf{1}} \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \hat{\mathbf{2}} \equiv \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \hat{\mathbf{3}} \equiv \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{56abc}$$

It then follows that

$$\begin{aligned} ^{(\text{vector})}H_{\bar{q}}^{*}(\bar{q}(-)) &= -2\{\bar{q}(-) \otimes \bar{\mathbf{v}}\}_R \{\bar{\mathbf{1}} | \bar{\mathbf{2}} | \bar{\mathbf{3}}\} \\ &= -2\{\bar{q}(-) \otimes \bar{\mathbf{v}}\}_R \begin{bmatrix} I_{3\times 3} \\ \mathbf{0}^{\mathrm{T}} \end{bmatrix} = -2\Xi(\bar{q}(-) \otimes \bar{\mathbf{v}}) \end{aligned} \tag{57}$$

Thus, equation (55) becomes

$$^{(\text{vector})}H_{\bar{q}}^{*}(\bar{q}(-)) = -2\Xi^{\mathrm{T}}(\bar{q}(-) \otimes \bar{\mathbf{v}}) \tag{58}$$

The matrix $\Xi(\bar{q})$, given by

$$\Xi(\bar{q}) = \begin{bmatrix} q_4 I_{3\times 3} - [\![\mathbf{q}]\!] \\ -\mathbf{q}^{\mathrm{T}} \end{bmatrix} \tag{59}$$

is familiar from the kinematic equation for the quaternion [3]

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{q} = \frac{1}{2}\Xi(\bar{q})\boldsymbol{\omega} \tag{60}$$

where $\boldsymbol{\omega}$ is the body-referenced angular velocity. For a more general measurement model, it suffices to replace $\mathbf{u}$ by $\nabla f(A(-)\mathbf{v})$ as before.

The Connection Between the Measurement Sensitivity Matrices

Equation (60) provides the first step of the proper path to obtaining a sensitivity matrix for the quaternion which embodies the norm constraint to first order. If we imagine the finite value of $\boldsymbol{\epsilon}$ as being due to a physical rotation, then the change in the quaternion related to the change in the error vector $\boldsymbol{\epsilon}$ is given by

$$\Delta\bar{q} = \frac{1}{2}\Xi(\bar{q})\boldsymbol{\epsilon} \tag{61a}$$

If $\Delta\bar{q}$ satisfies equation (61a), then form $\Xi^{\mathrm{T}}(\bar{q})\Xi(\bar{q}) = I_{3\times 3}$ it follows that

$$\boldsymbol{\epsilon} = 2\Xi^{\mathrm{T}}(\bar{q})\,\Delta\bar{q} \tag{61b}$$

The matrix $\Xi(\bar{q})$ has the property [3] that

$$\Xi^{\mathrm{T}}(\bar{q})\bar{q} = \mathbf{0} \tag{62}$$

Given an arbitrary quaternion, the matrix $\Xi^{\mathrm{T}}(\bar{q}(-))$ in equation (61b) will annihilate in $\Delta\bar{q}$ the component along $\bar{q}(-)$, which can affect the length of $\bar{q}(+)$ to first order. Thus, given a measurement sensitivity matrix for $\boldsymbol{\epsilon}$ we can construct the corresponding constraint-sensitive[15] measurement sensitivity matrix for $\Delta\bar{q}$ according to

$$H_{\bar{q}}(\bar{q}(-)) = 2H_{\epsilon}(A(\bar{q}(-)))\Xi^{\mathrm{T}}(\bar{q}(-)) \tag{63}$$

Such a measurement sensitivity matrix is not sensitive to the component of the quaternion along $\bar{q}(-)$. It is different from the quaternion measurement sensitivity matrices of the previous section. To distinguish these two the new quaternion measurement sensitivity matrix does not bear an asterisk.

We can now understand the argument of Ξ^{T} in equation (58). One could equally well compute $H^{*}_{\delta\bar{q}}$, in which case $\bar{q}(-) \rightarrow \delta\bar{q}(-) = \bar{1} = [0, 0, 0, 1]^{\mathrm{T}}$ (the identity quaternion) and $\mathbf{v} \equiv \mathbf{v}_I \rightarrow \mathbf{v}_B = A(-)\mathbf{v}$. Hence, the sensitivity matrix in the body frame becomes $-2\Xi^{\mathrm{T}}(\overline{A(-)\mathbf{v}})$, which is just the well known statement that the measurement is not sensitive to rotations about the line of sight of the observed object.

This transformation also works in the opposite direction.

$$H_{\epsilon}(A(\bar{q}(-))) = \frac{1}{2}H_{\bar{q}}(\bar{q}(-))\Xi(\bar{q}(-)) \tag{64}$$

$$H_{\epsilon}(A(\bar{q}(-))) = \frac{1}{2}H^{*}_{\bar{q}}(\bar{q}(-))\Xi(\bar{q}(-)) \tag{65}$$

which are no more than a statement of the chain rule for partial differentiation. Equation (65) can be tested by substituting equations (33) and (53) or equations (38) and (58). Finally

$$\begin{aligned} H_{\bar{q}}(\bar{q}(-)) &= 2H_{\epsilon}(A(\bar{q}(-)))\Xi^{\mathrm{T}}(\bar{q}(-)) \\ &= H^{*}_{\bar{q}}(\bar{q}(-))\Xi(\bar{q})\Xi^{\mathrm{T}}(\bar{q}) = H^{*}_{\bar{q}}(\bar{q}(-))\,(I - \bar{q}\,\bar{q}^{\mathrm{T}}) \end{aligned} \tag{66}$$

[15]"Constraint-sensitive" and "constraint-insensitive" in this paper will always refer to the quaternion norm or the orthogonality constraint of the attitude matrix. The measurement sensitivity matrix for $\boldsymbol{\epsilon}$ is always constraint-sensitive.

Note that all of these relationships assume that $\overline{q}(-)$ has unit norm.

The Sensitivity Matrix for an Arbitrary Attitude Representation

Equation (65) may be seen as an alternate method for computing $H_\epsilon(A(-))$. It is clear, however, from the lengthy derivation of the previous section that it is easier to compute $H_\epsilon(A(-))$ directly.

A similar result to equation (63) can be used to obtain the sensitivity matrix in terms of the Euler angles. The body-referenced angular velocity is related to the Euler angle rates by a relationship of the form [3]

$$\boldsymbol{\omega} = M(\varphi,\vartheta,\psi)\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\varphi\\ \vartheta\\ \psi\end{bmatrix} \equiv M(\boldsymbol{\phi})\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\phi} \tag{67}$$

This is equivalent to first order to writing [3] (in the notation of this work)

$$\boldsymbol{\epsilon}^* = M(\boldsymbol{\phi})\,(\boldsymbol{\phi}^* - \boldsymbol{\phi}^{\text{true}}) \tag{68}$$

and similarly for equations (71) and (72) below for the attitude correction [3].

By trivial inspection, the sensitivity matrix to changes in the Euler angles for a scalar measurement is

$$H_\phi(\boldsymbol{\phi}(-)) = H_\epsilon(A(\boldsymbol{\phi}(-)))M(\boldsymbol{\phi}(-)) \tag{69}$$

which, for a simple scalar measurement, becomes

$$H_\phi(\boldsymbol{\phi}(-)) = (\mathbf{u} \times A(\boldsymbol{\phi}(-))\mathbf{v})^{\mathrm{T}}M(\boldsymbol{\phi}(-)) \tag{70}$$

and similarly for vector measurements. Simple expressions [3, 17] exist for the matrix $M(\boldsymbol{\phi})$.

In general, if the inverse kinematic equation for a given attitude representation $\boldsymbol{\alpha}$ (of any dimension) is known, namely

$$\boldsymbol{\omega} = M(\boldsymbol{\alpha})\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\alpha} \tag{71}$$

then the corresponding constraint-sensitive measurement sensitivity matrix[16] is

$$H_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}(-)) = H_\epsilon(A(\boldsymbol{\alpha}(-)))M(\boldsymbol{\alpha}(-)) \tag{72}$$

Thus, despite our earlier assertion, it is a simple matter to construct the constraint-sensitive measurement sensitivity matrix for an arbitrary attitude representation provided that the inverse kinematic equation for the representation is known, but only because our predecessors have done the hard work to find an efficient form for the inverse kinematic equation.

Ambiguity of the Measurement Sensitivity Matrix for the Quaternion

The constraint-insensitive measurement sensitivity matrix for the quaternion depends on the *nonunique* relation of the attitude matrix to the quaternion. If in equation (43) the factor multiplying the identity matrix had been changed to $(1 - 2|\mathbf{q}|^2)$ to give

$$A(\overline{q}) = (1 - 2|\mathbf{q}|^2)I_{3\times3} + 2\mathbf{q}\mathbf{q}^{\mathrm{T}} + 2q_4[[\mathbf{q}]] \tag{73}$$

[16]Sensitivity to the constraint is an issue only if the dimension of $\boldsymbol{\alpha}$ is greater than three.

then the form of the sensitivity matrix would have been completely different. In particular, for a simple scalar measurement, the alternate parameterization, which is also popular though not as popular as equation (43), leads to the constraint-insensitive measurement sensitivity matrix

$$ {}^{(\mathrm{scalar})}H_{\bar{q}}^{*\prime}(\bar{q}(-)) = -2[\bar{\mathbf{u}} \otimes \bar{q}(-) \otimes \bar{\mathbf{v}}]^{\mathrm{T}} - 2(\mathbf{u}\cdot\mathbf{v})\bar{q}^{\mathrm{T}}(-) \tag{74} $$

A more general parameterization one can make for the attitude matrix as a function of the quaternion is

$$ A(\bar{q}) = A^{(h)}(\bar{q}) + (\bar{q}^{\mathrm{T}}\bar{q} - 1)Q(\bar{q}) \tag{75} $$

where $A^{(h)}(\bar{q})$ is the (homogeneous) parameterization of equation (43), and $Q(\bar{q})$ is an arbitrary differentiable 3×3 matrix function of the quaternion. For the alternate popular parameterization of equation (73), $Q(\bar{q})$ was simply $-I_{3\times3}$. $Q(\bar{q}) = +I_{3\times3}$ yields a variant of equation (73) with $(2q_4^2 - 1)$ multiplying the identity matrix. One could also multiply the right member of equation (43) by a 3×3 matrix function $T(\bar{q}^{\mathrm{T}}\bar{q})$ which is equal to the identity matrix for unit argument. The choice $T(x) = (1/x)I_{3\times3}$ would lead to the ray representation of the attitude.

Given equation (75), then, a more general constraint-insensitive measurement sensitivity matrix is

$$ {}^{(\mathrm{scalar})}H_{\bar{q}}^{*}(\bar{q}(-)) = {}^{(\mathrm{scalar})}H_{\bar{q}}^{*}(\bar{q}(-)) + \Delta({}^{(\mathrm{scalar})}H_{\bar{q}}^{*}(\bar{q}(-)) \tag{76a} $$

with

$$ \Delta({}^{(\mathrm{scalar})}H_{\bar{q}}^{*}(\bar{q}) = a(\bar{q})\bar{q}^{\mathrm{T}} + (\bar{q}^{\mathrm{T}}\bar{q} - 1)\mathbf{b}^{\mathrm{T}}(\bar{q}) \tag{76b} $$

and

$$ a(\bar{q}) = 2\mathbf{u}^{\mathrm{T}}Q(\bar{q})\mathbf{v}, \qquad \mathbf{b}^{\mathrm{T}}(\bar{q}) = \mathbf{u}^{\mathrm{T}}\frac{\partial Q}{\partial \bar{q}}(\bar{q})\mathbf{v} \tag{76cd} $$

For the vector measurement

$$ \Delta({}^{(\mathrm{vector})}H_{\bar{q}}^{*}(\bar{q})) = 2Q(\bar{q})\mathbf{v}\bar{q}^{\mathrm{T}} + (\bar{q}^{\mathrm{T}}\bar{q} - 1)\frac{\partial Q}{\partial \bar{q}}(\bar{q})\mathbf{v} \tag{77} $$

For $\bar{q}(-)$ having unit norm these measurement sensitivity matrices, when substituted into equation (65), will generate the same measurement sensitivity matrices for $\boldsymbol{\epsilon}$ independent of the value of the new terms. In estimation problems the constraint-insensitive quaternion measurement sensitivity matrices must be treated with extreme caution. Since the constraint-insensitive measurement sensitivity matrix, unlike the constraint-sensitive measurement sensitivity matrix, is not unique, there can be no simple transformation from the latter to the former in the manner of equation (63). Note that for some choices of $Q(\bar{q})$ the constraint-insensitive measurement sensitivity matrix will become sensitive to the constraint. Such a choice for $Q(\bar{q})$, in fact, leads to equation (74).

To appreciate the differences which can arise in the calculation of the constraint-insensitive measurement sensitivity matrices let us evaluate equations (53) and (74) for the case that $\bar{q}(-) = [0,0,0,1]^{\mathrm{T}}$. If we let $\mathbf{u} = [1,0,0]^{\mathrm{T}}$ and $\mathbf{v} = [\cos 30°, \sin 30°, 0]^{\mathrm{T}}$ we will arrive at the following numerical values for equations (53) and (74)

$$ H_{\bar{q}}^{*} = [0,0,1,1], \quad \text{and} \quad H_{\bar{q}}^{*\prime} = [0,0,1,0], \tag{78ab} $$

The differences are indeed substantial.

Despite this ambiguity in the constraint-insensitive measurement sensitivity matrices for the quaternion they can nonetheless be used in *constrained* quaternion estimation. The restriction of the quaternion $\bar{q}^*(+)$ to S^3 means that the first term in equation (76b) or equation (77) will not contribute at all because $\bar{q}^{\mathrm{T}}(-)\Xi(\bar{q}(-)) = \mathbf{0}^{\mathrm{T}}$, and likewise for the second term, because $\bar{q}(-)$ has unit norm. It is only *unconstrained* quaternion estimation that the ambiguity can appear to first order.

The Quaternion Norm and the Singularity of the Quaternion Covariance Matrix

The norm constraint on the quaternion leads necessarily to a nearly singular 4×4 quaternion covariance matrix when the rotations are confined to some minute region of $SO(3)$ (for quaternions some minute region (or, better, two minute antipodal regions) of S^3, the sphere in four dimensions. To see this consider the following probability distribution of the quaternion

$$\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{3\times3}), \qquad q_4 = \sqrt{1 - |\mathbf{q}|^2} \tag{79ab}$$

We assume that $\sigma^2 \ll 1$. This normal distribution for $\mathbf{q}$ is, of course, not rigorously possible because it allows $|\mathbf{q}| > 1$, but the probability of that occurring is so infinitesimal that we can ignore it, provided we are careful in our treatment to truncate the integrals so as not to allow the square roots to have (infinitesimally) negative arguments. From the consideration of symmetry it is obvious that the covariance matrix of the quaternion must have the form[17]

$$P_{\bar{q}\bar{q}} = \mathrm{diag}(\sigma^2, \sigma^2, \sigma^2, \tau^2) \tag{80}$$

where the right member denotes a diagonal matrix labeled by its diagonal elements. Straightforward calculation leads to[18]

$$\tau^2 = \frac{3}{2}\sigma^4 + O(\sigma^6) \tag{81}$$

The 4×4 covariance matrix is not strictly singular, but it is certainly very ill conditioned. For $\sigma = 2$ arc sec, a routinely achievable attitude accuracy today, the condition number (the ratio of the largest to the smallest eigenvalue of a matrix) is greater than 10^9, which makes the matrix singular for all practical purposes. This statement, clearly, has nothing to do with the sensor models or with the estimation method.

Note also that the mean quaternion in our examples is given by

$$\bar{\mu}_{\bar{q}} = \left[0, 0, 0, \sqrt{1 - 3\sigma^2}\right]^{\mathrm{T}} + O(\sigma^4) \tag{82}$$

which is not a possible quaternion of rotation, once more demonstrating that the mean and covariance matrix are not rigorous statistical concepts for random vectors defined on curved surfaces.

[17]The result is not changed if the sign of the quaternion is allowed to vary.
[18]The second moment of $q_4 - 1$ is $(15/4)\sigma^2$. Note that the distribution of q_4 is not Gaussian but approximately $1 + (\sigma^2/2)\chi^2(3)$.

The most general result for $\overline{\mu}_{\overline{q}}$ and $P_{\overline{q}\,\overline{q}}$ is

$$\overline{\mu}_{\overline{q}} = \{\overline{q}\}_R\left[0, 0, 0, \sqrt{1 - \operatorname{tr} P_{\delta\mathbf{q}\delta\mathbf{q}}}\right]^{\mathrm{T}} + O(|\operatorname{tr} P_{\delta\mathbf{q}\delta\mathbf{q}}|^2) \tag{83a}$$

$$P_{\overline{q}\,\overline{q}} = \{\overline{q}\}_R \begin{bmatrix} P_{\delta\mathbf{q}\delta\mathbf{q}} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & P_{\delta q_4 \delta q_4} \end{bmatrix} \{\overline{q}\}_R^{\mathrm{T}} \tag{83b}$$

with $P_{\delta\mathbf{q}\delta\mathbf{q}} = P_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}/4$ and

$$P_{\delta q_4 \delta q_4} = \frac{1}{4}(\operatorname{tr} P_{\delta\mathbf{q}\delta\mathbf{q}})^2 + (\operatorname{tr}\,(P_{\delta\mathbf{q}\delta\mathbf{q}})^2) + O((\operatorname{tr} P_{\delta\mathbf{q}\delta\mathbf{q}})^3) \tag{83c}$$

This is not the form that would be obtained from an estimation procedure, in which $\Delta(\delta q_4)$ would vanish identically within the linearized approximation. Thus, the 4×4 quaternion covariance matrix arising from a typical iterative estimation problem would be

$$P_{\overline{q}\,\overline{q}} = \{\overline{q}\}_R \begin{bmatrix} P_{\delta\mathbf{q}\delta\mathbf{q}} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 0 \end{bmatrix} \{\overline{q}\}_R^{\mathrm{T}} = \Xi(\overline{q}) P_{\delta\mathbf{q}\delta\mathbf{q}} \Xi^{\mathrm{T}}(\overline{q}) \tag{84}$$

In computations in which terms of order $|\operatorname{tr} P_{\delta\mathbf{q}\delta\mathbf{q}}|^2$ are discarded, $P_{\overline{q}\,\overline{q}}$ can certainly be said to be singular. We shall see in Part II that the 4×4 covariance matrix of equation (84) is more appropriate than that of equation (83).

It has sometimes been said that because $\Delta\overline{q}$ is not a quaternion of rotation it satisfies no constraint. That is obviously not true. As stated at the end of Part I, since $\overline{q}(+)$ must have unit norm, it follows that

$$|\overline{q}(+)|^2 - 1 = 2\overline{q}^{\mathrm{T}}(-)\Delta\overline{q}(+) + |\Delta\overline{q}(+)|^2 = 0 \tag{85}$$

so that the component of $\Delta\overline{q}(+)$ along $\overline{q}(-)$ must nearly vanish.

If we denote by $\Delta q_\parallel$ the component of $\Delta\overline{q}(+)$ in the direction of $\overline{q}(-)$ and by $\Delta\mathbf{q}_\perp$ the three components of $\Delta\overline{q}(+)$ in the tangent space to $\overline{q}(-)$, then these must satisfy

$$(\Delta q_\parallel)^2 + 2\Delta q_\parallel + |\Delta\mathbf{q}_\perp|^2 = 0 \tag{86}$$

which is readily solved to give

$$\Delta q_\parallel = -1 + \sqrt{1 - |\Delta\mathbf{q}_\perp|^2} = -|\Delta\mathbf{q}_\perp|^2/2 + O(|\Delta\mathbf{q}_\perp|^4) \tag{87}$$

$\Delta\overline{q}(+)$ is most certainly constrained. Not surprisingly, the relationship between $\Delta q_\parallel$ and $\Delta\mathbf{q}_\perp$ is identical to that between δq_4 and $\delta\mathbf{q}$. Clearly, $\Delta q_\parallel$ should not be treated as a free variable independent of $\Delta\mathbf{q}_\perp$. In fact, it is easy to show that $\Delta q_\parallel$ must be $\delta q_4 - 1$ and that $\Delta\mathbf{q}_\perp$ can be chosen to be $\delta\mathbf{q}$. (See Appendix B.)

The near-singularity of the 4×4 quaternion covariance matrix depends, of course, on the attitude being confined to a small region (or equivalently to two small antipodal regions). Even with measurement accuracies of 0.5 deg we anticipate 4×4 quaternion covariance matrices with condition numbers of about 100. If, however, we consider the case of an attitude which is completely unknown, then the probability distribution of the quaternion will be "uniform" over the entire quaternion space. The uniform probability density function for the vector components of the quaternion has been shown to be [18, 19]

$$p_q(\mathbf{q}') = \frac{1}{\pi^2\sqrt{1 - |\mathbf{q}'|^2}}, \qquad 0 \leq |\mathbf{q}'| < 1 \tag{88}$$

with q_4 a function of $\mathbf{q}$. In the four-dimensional space of the quaternion this becomes[19]

$$p_{\bar{q}}(\bar{q}') = \frac{1}{2\pi^2}\,\delta(q' - 1) \tag{89}$$

where $\delta(x)$ is the Dirac δ-function and q is the value of the quaternion magnitude. The quaternion is clearly uniformly distributed over a spherical surface in four dimensions. The quaternion statistics in this case are

$$\bar{\mu}_{\bar{q}} = \bar{0} \quad \text{and} \quad P_{\bar{q}\bar{q}} = \frac{1}{4} I_{4\times 4} \tag{90ab}$$

The quaternion 4×4 covariance is certainly non-singular in this instance. However, the mean quaternion is $\bar{0} = [0,0,0,0]^{\mathrm{T}}$, which is highly unphysical and so far from a physical solution that were this used as an initial condition for an estimation procedure, a linearization about the initial mean would be meaningless. The uniform distribution of $\mathbf{q}$ is also not Gaussian.

To a reasonable person these remarks about the 4×4 quaternion covariance matrix must seem, well, reasonable, and to some degree they are. Nonetheless, we shall review our statements here very harshly at the end of Part II.

The QUEST Measurement Model, the Wahba Cost Function, and MLE

The QUEST measurement model [5, 20] was developed for a brute-force calculation of the estimate error covariance matrix for the Wahba problem, that is to develop an approximate estimate error covariance matrix for the attitude matrix estimate which minimizes the Wahba cost function [20, 21], which, at the time, was thought to have no rigorous connection to MLE. It was shown later that within MLE the QUEST measurement model led directly to the Wahba problem [5]. However, this is not true if the attitude matrix is not constrained to be proper orthogonal. In the present section we will develop the QUEST measurement model and derive the Wahba problem from it. We shall also present a cost function which is consistent with the QUEST measurement model and a measurement model which is consistent with the Wahba cost function.

The general model for unit-vector measurements is

$$\mathbf{z}_k = A\hat{\mathbf{V}}_k + \boldsymbol{\eta}_k \tag{91}$$

or in more familiar notation

$$\hat{\mathbf{W}}_k = A\hat{\mathbf{V}}_k + \Delta\hat{\mathbf{W}}_k \quad \text{with} \quad \Delta\hat{\mathbf{W}}_k \sim \mathcal{N}(\mathbf{0}, R_k) \tag{92ab}$$

Since $\hat{\mathbf{W}}_k$ is a unit vector, there can be no error to first order in $\hat{\mathbf{W}}_k$ in the direction of $\hat{\mathbf{W}}_k^{\text{true}}$. Therefore, it must be true that

$$\hat{\mathbf{W}}_k^{\text{true}\,\mathrm{T}}\Delta\hat{\mathbf{W}}_k = O(\operatorname{tr} R_k) \tag{93}$$

It follows that

$$R_k\hat{\mathbf{W}}_k^{\text{true}} = O((\operatorname{tr} R_k)^{3/2}) \tag{94}$$

[19]Note that equation (88) assumes that $\bar{\mathbf{q}}'$ is confined to the hyperhemisphere (hemihypersphere?) $q_4' > 0$, while equation (89) assumes that it ranges over the entire hypersphere S^3. Note that in equations (88) and (89) we distinguish the sampled value from the random variable by a prime.

Hence, there must exist an orthonormal triad $\{\hat{\mathbf{W}}_k^{\text{true}}, \hat{\mathbf{U}}_{k,1}, \hat{\mathbf{U}}_{k,2}\}$ such that

$$R_k = \sigma_{k,1}^2 \hat{\mathbf{U}}_{k,1}\hat{\mathbf{U}}_{k,1}^{\mathrm{T}} + \sigma_{k,2}^2 \hat{\mathbf{U}}_{k,2}\hat{\mathbf{U}}_{k,2}^{\mathrm{T}} + O((\operatorname{tr} R_k)^{3/2}) \tag{95}$$

The QUEST measurement model (QMM) consists of setting

$$\sigma_{k,1}^2 = \sigma_{k,2}^2 = \sigma_k^2 \tag{96}$$

and ignoring the higher-order terms, whence

$$R_k = \sigma_k^2 (I_{3\times 3} - \hat{\mathbf{W}}_k^{\text{true}} \hat{\mathbf{W}}_k^{\text{true}\,\mathrm{T}}) \tag{97}$$

The QMM measurement covariance matrix is singular, hence not invertible. To construct the MLE cost function we consider the two scalar measurements

$$z_k(j) \equiv \hat{\mathbf{U}}_{k,j}^{\mathrm{T}} \hat{\mathbf{W}}_k = \hat{\mathbf{U}}_{k,j}^{\mathrm{T}} A \hat{\mathbf{V}}_k + \eta_{k,j}, \quad k = 1, \ldots, N, \quad j = 1, 2 \tag{98}$$

with $\eta_{k,j} \sim \mathcal{N}(0, \sigma_k^2)$. By definition, $z_k(j)$ will always have a sampled value close to zero, but this is of no consequence. The MLE result for the QMM cost function is

$$\begin{aligned} J(A) &= \sum_{k=1}^{N} \sum_{j=1}^{2} \frac{1}{\sigma_k^2} |\hat{\mathbf{U}}_{k,j}^{\mathrm{T}} (\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k)|^2 \\ &= \sum_{k=1}^{N} \frac{1}{\sigma_k^2} (\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k)^{\mathrm{T}} (\hat{\mathbf{U}}_{k,1}\hat{\mathbf{U}}_{k,1}^{\mathrm{T}} + \hat{\mathbf{U}}_{k,2}\hat{\mathbf{U}}_{k,2}^{\mathrm{T}})(\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k) \\ &= \sum_{k=1}^{N} \frac{1}{\sigma_k^2} (\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k)^{\mathrm{T}} (I - \hat{\mathbf{W}}_k^{\text{true}} \hat{\mathbf{W}}_k^{\text{true}\,\mathrm{T}})(\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k) \end{aligned} \tag{99}$$

Thus, the effective inverse covariance matrix for the correct cost function for batch unconstrained attitude matrix estimation from direction measurements is

$$\text{"}R_k^{-1}\text{"} = \frac{1}{\sigma_k^2} (I - \hat{\mathbf{W}}_k^{\text{true}} \hat{\mathbf{W}}_k^{\text{true}\,\mathrm{T}}) \tag{100}$$

which is the pseudo-inverse of the QUEST model measurement covariance matrix. We have now developed the QUEST measurement model and derived the resulting QMM cost function. We note that the QMM cost function is not that of Wahba.

Consider now the iterative minimization of the QMM cost function over the manifold $SO(3)$. At the i-th iteration we have for the QMM cost function

$$\begin{aligned} J^{\text{QMM}}(\boldsymbol{\epsilon}_i) = \frac{1}{2} \sum_{k=1}^{N} \frac{1}{\sigma_k^2} [\hat{\mathbf{W}}_k - A_i(-)\hat{\mathbf{V}}_k + [\![A_i(-)\hat{\mathbf{V}}_k]\!]\boldsymbol{\epsilon}_i]^{\mathrm{T}} \\ (I - \hat{\mathbf{W}}_k^{\text{true}} \hat{\mathbf{W}}_k^{\text{true}\,\mathrm{T}}) [\hat{\mathbf{W}}_k - A_i(-)\hat{\mathbf{V}}_k + [\![A_i(-)\hat{\mathbf{V}}_k]\!]\boldsymbol{\epsilon}_i] \end{aligned} \tag{101}$$

whence

$$\begin{aligned} \frac{\partial J^{\text{QMM}}}{\partial \boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_i^*(+)) = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} [\![A_i(-)\hat{\mathbf{V}}_k]\!]^{\mathrm{T}} (I - \hat{\mathbf{W}}_k^{\text{true}} \hat{\mathbf{W}}_k^{\text{true}\,\mathrm{T}}) \\ [\hat{\mathbf{W}}_k - A_i(-)\hat{\mathbf{V}}_k + [\![A_i(-)\hat{\mathbf{V}}_k]\!]\boldsymbol{\epsilon}_i^*(+)] = \mathbf{0} \end{aligned} \tag{102}$$

We now note that asymptotically (i.e., as $N \rightarrow \infty$)

$$\lim_{i \to \infty} [\![A_i(-)\hat{\mathbf{V}}_k]\!] = [\![\hat{\mathbf{W}}_k^{\text{true}}]\!] \tag{103}$$

Thus, asymptotically, the term in R_k^{-1} proportional to $\hat{\mathbf{W}}_k^{\text{true}}\hat{\mathbf{W}}_k^{\text{true}\,T}$ does not contribute in the asymptotic limit when the iterative estimation procedure has converged. The resulting cost function when this term is discarded is

$$J^{\text{Wahba}}(A) \equiv \sum_{k=1}^{N}\sum_{j=1}^{2}\frac{1}{\sigma_k^2}|\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k|^2 \tag{104}$$

which is the Wahba cost function. The substitution of $J^{\text{Wahba}}(A)$ for $J^{\text{QMM}}(A)$ required that A be restricted to the $SO(3)$ manifold and the asymptotic approximation.

If we define

$$\frac{1}{\sigma_{\text{tot}}^2} \equiv \sum_{k=1}^{N}\frac{1}{\sigma_k^2} \tag{105}$$

and

$$a_k \equiv \sigma_{\text{tot}}^2/\sigma_k^2 \quad \text{so that} \quad \sum_{k=1}^{N} a_k = 1 \tag{106ab}$$

then we can rewrite the Wahba cost function as

$$J^{\text{Wahba}}(A) = \frac{1}{2\sigma_{\text{tot}}^2}\sum_{k=1}^{N} a_k|\hat{\mathbf{W}}_k - A\hat{\mathbf{V}}_k|^2 \tag{107}$$

a more familiar form, except for the preceding factor $1/\sigma_{\text{tot}}^2$, a requirement of MLE.[20]

It is now a trivial matter to present the measurement model which leads via MLE to the Wahba cost function. That is, namely,

$$\mathbf{W}_k = A\mathbf{V}_k + \Delta\mathbf{W}_k \quad \text{with} \quad \Delta\mathbf{W}_k \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 I_{3\times 3}) \tag{108ab}$$

We call this the *full-vector* measurement model. It is a possible model to use for the three-axis magnetometer if we define

$$\mathbf{z}_k = \mathbf{B}_k + \boldsymbol{\eta}_k \quad \text{with} \quad \boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 I_{3\times 3}) \tag{109ab}$$

(Note that $\mathbf{W}_k$ and σ_k^2 are no longer unitless.) Unfortunately, there exists only one sensor which could be correctly modeled[21] as a full-vector sensor. To use the full-vector model for a focal-plane sensor, such as a vector Sun sensor or a star camera, would be unphysical, to say the least. The detailed examination of these three models will be carried out in Part II [1].

We now have three models for studies of unconstrained attitude estimation: (1) The QUEST measurement model with its associated QMM cost function; (2) the full-vector measurement model with its associated Wahba cost function; and (3) a hybrid model consisting of the QUEST measurement model but the Wahba cost function.[22] We could, of course, consider mixtures of measurements which are either unit vectors obeying the QUEST measurement model or three-axis magnetometer measurements obeying the full-vector measurement model, but the three

[20]The normalization of the Wahba cost function in equation (104) leads to values of λ_{max} in QUEST which are approximately $1/\sigma_{\text{tot}}^2$ [5]. If $1/\sigma_k^2$ is replaced by a_k of equation (106b), as in the original publication of QUEST [20], then $\lambda_{\text{max}} \approx 1$. Obviously, the value of the QUEST quaternion is the same in both cases.

[21]Correctly modeled in the qualitative sense only, because the effective measurement covariance matrix will be dominated by errors in the reference magnetic field which are not isotropic.

[22]Obviously, there is no point in considering the QMM cost function and the full-vector measurement model. Combination (3), at least, has a precedent.

above will be sufficient for our investigatory purposes and more transparent. Note that only the first two models are consistent with MLE if the constraint is relaxed and these are the only models which will have significance for the Kalman filter.

A Failure of Unconstrained Quaternion Estimation

Consider the three *noise-free* measurements

$$\mathbf{z}_1 = A(\bar{q}_{\text{UC}})\hat{\mathbf{1}} = \hat{\mathbf{1}}, \quad \mathbf{z}_2 = A(\bar{q}_{\text{UC}})\hat{\mathbf{2}} = \hat{\mathbf{2}}, \quad \mathbf{z}_3 = A(\bar{q}_{\text{UC}})\hat{\mathbf{3}} = \hat{\mathbf{3}} \tag{110abc}$$

Clearly, by linear superposition $A(\bar{q})\mathbf{V} = \mathbf{V}$ for any 3×1 column vector $\mathbf{V}$. Hence, to add further noise-free measurements would be superfluous (in the constrained estimation case, equation (110c) would also be superfluous), and equations (110) are identical to $A(\bar{q}_{\text{UC}}) = I_{3\times3}$. It follows that we must solve for the quaternion from the nine equations

$$A(\bar{q}_{\text{UC}}) = I_{3\times3} \tag{111}$$

Taking the attitude matrix-quaternion relation to be given by equation (43) we obtain directly

$$A - A^{\text{T}} = 4q_{4\text{UC}}[\![\mathbf{q}_{\text{UC}}]\!] = \mathbf{0}_{3\times3} \tag{112}$$

whence, either $q_{4\text{UC}} = 0$ or $\mathbf{q}_{\text{UC}} = \mathbf{0}$. However

$$\text{tr}\, A = 3(q_{4\text{UC}})^2 - |\mathbf{q}_{\text{UC}}|^2 = 3 \tag{113}$$

for which $q_{4\text{UC}} = 0$ is impossible. Hence

$$\mathbf{q}_{\text{UC}} = \mathbf{0}, \quad q_{4\text{UC}} = \pm 1 \tag{114ab}$$

which are the correct values for the quaternion.

If, on the other hand, we choose equation (73) as the relationship between the attitude matrix and the quaternion, we must solve instead

$$A_{ij}(\bar{q}_{\text{UC}}) = \begin{cases} (1 - 2|\mathbf{q}|^2 + 2q_i^2)_{\text{UC}} = 1 & \text{for } i = j \\ \left(2q_iq_j + q_4\sum_{k=1}^{3}\epsilon_{ijk}q_k\right)_{\text{UC}} = 0 & \text{for } i \neq j \end{cases} \tag{115}$$

where ϵ_{ijk} is the Levi-Civita symbol [3]. For $i = j$ we have directly

$$(q_1^2 + q_2^2)_{\text{UC}} = (q_1^2 + q_3^2)_{\text{UC}} = (q_2^2 + q_3^2)_{\text{UC}} = 0 \tag{116}$$

whence

$$q_{1\text{UC}} = q_{2\text{UC}} = q_{3\text{UC}} = 0 \tag{117}$$

The remaining equations all become

$$0 + q_{4\text{UC}} \cdot 0 = 0 \tag{118}$$

We see that $q_{4\text{UC}}$ is indeterminate in this case. This will not remain the situation when noise is added to the measurements, but the estimate for $q_{4\text{UC}}$ under these circumstances cannot be physically meaningful. Thus, for a very reasonable situation the unconstrained estimation of the quaternion can yield very stupid results. This will be observed again in the more realistic examples of Part II [1].

Note that there is no similar problem for the estimation of the unconstrained attitude matrix, which yields directly $A^*_{\text{UC}} = I_{3\times3}$.

Summary and Discussion

We have given a very complete and careful presentation of the foundations of batch least-squares attitude estimation and the attitude measurement sensitivity matrix. We have shown, in particular, that the constraint-insensitive quaternion measurement sensitivity matrices are ambiguous, hence, unconstrained quaternion estimation will lead to ambiguous, and hence meaningless, estimates for the quaternion. We gave a specific example in which, effectively, one choice of the measurement sensitivity matrix led to the correct estimate of the quaternion for noise-free measurements, while another choice did not lead to an estimate at all. Further, we showed that the Wahba problem results from the QUEST measurement model, a very physical model for direction measurements, only when the attitude is constrained. When this is not the case the least-squares cost function must have a different form. Thus, we see from the outset that unconstrained quaternion estimation is beset with problems.

Since, the quaternion measurement sensitivity matrix is ambiguous, is there a criterion for selecting the right one? The answer is simply that since the unconstrained quaternion estimate is also meaningless, the best quaternion measurement sensitivity matrix to use is the one which leads eventually to the most efficient calculation of the constrained estimate of the quaternion. By this we mean not any quaternion estimate which would have had unit norm but the quaternion estimate that would have been obtained had we insisted on maintaining the constraint correctly throughout the estimation process. This turns out to be a tall order, as we shall see in Part II of this work [1].

Acknowledgments

The author is grateful to Itzhack Bar-Itzhack, Yang Cheng, Richard Harman, F. Landis Markley, Yaakov Oshman, Mark Pittelkau, Mark Psiaki, and Julie Thienel (formerly Deutschmann) for many interesting discussions on this topic over the years. Dr. Markley is to especially thanked for his very critical reading on the manuscript and his pointing out of numerous errors of presentation.

References

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

[2] SHUSTER, M. D. "The Quaternion in the Kalman Filter," Paper No. AAS-93–62, *Proceedings, AAS/AIAA Astrodynamics Specialists Conference*, Victoria, British Columbia, Canada, August 1993; printed as "The Quaternion in the Kalman Filter," *Advances in the Astronautical Sciences*, Vol. 85, pp. 25–37, 1993.

[3] SHUSTER, M. D. "A Survey of Attitude Representations," *The Journal of the Astronautical Sciences*, Vol. 41, No. 4, pp. 439–517, October–December 1993.

[4] SORENSON, H. W. *Parameter Estimation*, Marcel Dekker, New York, 1980.

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

[6] FALLON, L. III and RIGTERINK, P. V. "Introduction to Estimation Theory," in WERTZ, J. R. (ed.), *Spacecraft Attitude Determination and Control*, Kluwer Academic, Dordrecht, 1978, pp. 447–459.

[7] SHUSTER, M. D., PITONE, D. S. and BIERMAN, G. J. "Batch Estimation of Spacecraft Sensor Alignments," *The Journal of the Astronautical Sciences*, Vol. 39, No. 4, October–December 1991, Part I: pp. 519–546, Part II: pp. 547–571.

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

[9] SHUSTER, M. D. "Erratum: Kalman Filtering of Spacecraft Attitude and the QUEST Model," *The Journal of the Astronautical Sciences*, Vol. 51, No. 3, July–September 2003.
[10] MARKLEY, F. L. "Attitude Error Representations for Kalman Filtering," to appear, *Journal of Guidance, Control and Dynamics*.
[11] 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, September–October 1982, pp. 417–429.
[12] BAR-ITZHACK, I. Y. and FEGLEY, K. A. "Orthogonalization Techniques of a Direction Cosine Matrix," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-5, September 1969, pp. 798–804.
[13] BAR-ITZHACK, I. Y. "Iterative Optimal Orthogonalization of the Strapdown Matrix," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-11, January 1975, pp. 30–37.
[14] BAR-ITZHACK, I. Y., MEYER, J., and FUHRMANN, P. A. "Strapdown Matrix Orthogonalization: the Dual Iterative Algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-12, January 1976, pp. 32–37.
[15] BAR-ITZHACK, I. Y. and MEYER, J. "On the Convergence of Iterative Orthogonalization Processes," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-12, March 1976, pp. 146–151.
[16] MEYER, J. and BAR-ITZHACK, I. Y. "Practical Comparison of Iterative Matrix Orthogonalization Algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-13, May 1977, pp. 230–235.
[17] BROUCKE, R. A. "Equations of Motion of a Rotating Rigid Body," *Journal of Guidance, Control and Dynamics*, Vol. 13, 1990, pp. 1150–1152.
[18] SHUSTER, M. D. "Man, Like These Attitudes Are Totally Random!" *Advances in the Astronautical Sciences*, Vol. 108, December 2001, Part I: pp. 383–396, Part II: pp. 397–408.
[19] SHUSTER, M. D. "Uniform Attitude Probability Distributions," to appear, *The Journal of the Astronautical Sciences*, Vol. 51, No. 4, October–December 2003.
[20] SHUSTER, M. D. and OH, S. D. "Three-Axis Attitude Determination from Vector Observations," *Journal of Guidance, Control and Dynamics*, Vol. 4, No. 1, Jan.–Feb. 1981 pp. 70–77.
[21] WAHBA, G. "A Least-Squares Estimate of Satellite Attitude," Problem 65–1, *SIAM Review*, Vol. 7, No. 3, July 1965, p. 409.

Appendix A: Transformation of Focal-Plane Measurements

The focal-plane measurements of equations (29) have the disadvantage, unlike the scalar and vector measurements, of not being linear in the attitude matrix. Generally, it is much easier to deal with measurements which are linear functions of the attitude matrix rather than with rational functions. In fact, with the exception of the three-axis magnetometer, our vector measurements all start out as focal-plane measurements.

To avoid a messy notation, let us write $\hat{\mathbf{W}} = [W_1, W_2, W_3]$ as the measured unit vector. Then we can write the focal-plane measurement vector as

$$\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} = \begin{bmatrix} W_1/W_3 \\ W_2/W_3 \end{bmatrix} \equiv \mathbf{f}(\hat{\mathbf{W}}), \quad \hat{\mathbf{W}} = \frac{1}{\sqrt{1 + \zeta_1^2 + \zeta_2^2}} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ 1 \end{bmatrix} \tag{A1ab}$$

A small variation $\Delta\hat{\mathbf{W}}$ is equivalent to a variation $\Delta\boldsymbol{\zeta}$ according to

$$\Delta\boldsymbol{\zeta} = \frac{1}{W_3} \begin{bmatrix} \Delta W_1 - \zeta_1 \Delta W_3 \\ \Delta W_2 - \zeta_2 \Delta W_3 \end{bmatrix} = U \Delta\hat{\mathbf{W}} \tag{A2}$$

with

$$U = \frac{1}{W_3} \begin{bmatrix} 1 & 0 & -\zeta_1 \\ 0 & 1 & -\zeta_2 \end{bmatrix} = \frac{1}{W_3} [I_{2\times 2} \,\vdots\, -\boldsymbol{\zeta}] \tag{A3}$$

Equation (A2) and the QUEST measurement model lead to a 2×2 measurement covariance matrix which is [6, 9]

$$R_\zeta = \sigma^2(1 + \zeta_1^2 + \zeta_2^2)\begin{bmatrix} 1 + \zeta_1^2 & \zeta_1\zeta_2 \\ \zeta_2\zeta_1 & 1 + \zeta_2^2 \end{bmatrix} \tag{A4}$$

Beginning instead with a known R_ξ we can obtain $R_{\hat{\mathbf{W}}}$ knowing that the three-dimensional measurement equation must have the form

$$\hat{\mathbf{W}} = A\hat{\mathbf{V}} + \boldsymbol{\eta}_{3\times1} \quad \text{with} \quad \boldsymbol{\eta}_{3\times1} \sim \mathcal{N}(\mathbf{0}, R_{\hat{\mathbf{W}}}) \tag{A5}$$

What we must derive is the form for the singular $R_{\hat{\mathbf{W}}}^{-1}$ with which to construct the cost function. Note that $R_{\hat{\mathbf{W}}}$ and $R_{\hat{\mathbf{W}}}^{-1}$ are both singular and therefore not related by inversion. By $R_{\hat{\mathbf{W}}}^{-1}$ we really mean the (singular) Fisher information associated with the measurement.

The contribution of the measurement $\boldsymbol{\epsilon}$ to the least-squares cost function is

$$\begin{aligned} J(A) &= \frac{1}{2}[\boldsymbol{\zeta} - \mathbf{f}(A\hat{\mathbf{V}})]^{\mathrm{T}} R_\zeta^{-1}[\boldsymbol{\zeta} - \mathbf{f}(A\hat{\mathbf{V}})] \\ &= \frac{1}{2}[\boldsymbol{\zeta} - \mathbf{f}(A(-)\hat{\mathbf{V}} + [\![\boldsymbol{\epsilon}]\!]A(-)\hat{\mathbf{V}})]^{\mathrm{T}} R_\zeta^{-1}[\boldsymbol{\zeta} - \mathbf{f}(A(-)\hat{\mathbf{V}} + [\![\boldsymbol{\epsilon}]\!]A(-)\hat{\mathbf{V}})] \\ &= \frac{1}{2}[\boldsymbol{\zeta} - \mathbf{f}(A(-)\hat{\mathbf{V}}) - U[\![\boldsymbol{\epsilon}]\!]A(-)\hat{\mathbf{V}}]^{\mathrm{T}} R_\zeta^{-1}[\boldsymbol{\zeta} - \mathbf{f}(A(-)\hat{\mathbf{V}}) - U[\![\boldsymbol{\epsilon}]\!]A(-)\hat{\mathbf{V}}] \\ &= \frac{1}{2}[\boldsymbol{\zeta} - \mathbf{f}(A(-)\hat{\mathbf{V}}) + U[\![A(-)\hat{\mathbf{V}}]\!]\boldsymbol{\epsilon}]^{\mathrm{T}} R_\zeta^{-1}[\boldsymbol{\zeta} - \mathbf{f}(A(-)\hat{\mathbf{V}}) + U[\![A(-)\hat{\mathbf{V}})]\!]\boldsymbol{\epsilon}] \\ &\equiv J(\boldsymbol{\epsilon}) \end{aligned} \tag{A6}$$

The contribution of this term to the Fisher information is

$$\frac{\partial^2 J(\boldsymbol{\epsilon})}{\partial\boldsymbol{\epsilon}^{\mathrm{T}}\partial\boldsymbol{\epsilon}} = [\![A(-)\hat{\mathbf{V}}]\!]^{\mathrm{T}} U^{\mathrm{T}} R_\zeta^{-1} U [\![A(-)\hat{\mathbf{V}}]\!] \tag{A7}$$

which is the same Fisher information that would have been obtained from the least-squares cost function

$$J(A) = \frac{1}{2}[\hat{\mathbf{W}} - A\hat{\mathbf{V}}]^{\mathrm{T}} U^{\mathrm{T}} R_\zeta^{-1} U[\hat{\mathbf{W}} - A\hat{\mathbf{V}}] \tag{A8}$$

Hence,

$$R_{\hat{\mathbf{W}}}^{-1} = U^{\mathrm{T}} R_\zeta^{-1} U \tag{A9}$$

is the desired inverse covariance matrix for constructing the weighted least-squares cost function for the equivalent vector measurement.

Of course, working directly from the QUEST measurement model (see equation (99)) we have

$$R_{\hat{\mathbf{W}}}^{-1} = \frac{1}{\sigma_k^2}(I_{3\times3} - \hat{\mathbf{W}}^{\mathrm{true}}\hat{\mathbf{W}}^{\mathrm{true\,T}}) \tag{A10}$$

Appendix B: Decomposition of the Quaternion Space

From

$$\{\bar{q}\}_R = [\Xi(\bar{q}) \vdots \bar{q}] \tag{B1}$$

it follows that

$$\Xi(\bar{q}(-))\Xi^{\mathrm{T}}(\bar{q}(-)) + \bar{q}(-)\bar{q}^{\mathrm{T}}(-) = I_{4\times 4} \tag{B2}$$

provided $\bar{q}(-)$ has unit norm. Likewise if $\bar{q}(-)$ has unit norm, we may write for any quaternion $\bar{p}$ of arbitrary norm

$$\begin{aligned}\bar{p} &= \Xi(\bar{q}(-))\mathbf{p}_\perp + \bar{q}(-)p_\parallel, \\ &\equiv \bar{p}_\perp + \bar{p}_\parallel\end{aligned} \tag{B3}$$

with

$$\mathbf{p}_\perp \equiv \Xi^{\mathrm{T}}(\bar{q}(-))\bar{p}, \quad p_\parallel \equiv \bar{q}^{\mathrm{T}}(-)\bar{p} \tag{B4ab}$$

Likewise, we may decompose an a 4×4 matrix M as

$$\begin{aligned}M = {} & \Xi(\bar{q}(-))M_{\perp\perp}\Xi^{\mathrm{T}}(\bar{q}(-)) + \bar{q}(-)M_{\parallel\parallel}\bar{q}^{\mathrm{T}}(-) \\ & + \Xi(\bar{q}(-))M_{\perp\parallel}\bar{q}^{\mathrm{T}}(-) + \bar{q}(-)M_{\parallel\perp}\Xi^{\mathrm{T}}(\bar{q}(-))\end{aligned} \tag{B5}$$

with $M_{\perp\perp} = \Xi^{\mathrm{T}}(\bar{q}(-))M\Xi(\bar{q}(-))$, etc.

For $\bar{p} = \bar{q}$, the quaternion in the constrained quaternion Kalman filter,

$$q_\parallel = \delta q_4, \quad \mathbf{q}_\perp = \delta\mathbf{q} \tag{B6}$$