AAS 10-235

A MULTIPLICATIVE RESIDUAL APPROACH TO ATTITUDE KALMAN FILTERING WITH UNIT-VECTOR MEASUREMENTS

Renato Zanetti*

Using direction vectors of unit length as measurements for attitude estimation in an extended Kalman filter inevitably results in a singular measurement covariance matrix. Singularity of the measurement covariance means no noise is present in one component of the measurement. Singular measurement covariances can be dealt with by the classic Kalman filter formulation as long as the estimated measurement covariance is non singular in the same direction. Unit vector measurements violate this condition since both the true measurement and the estimated measurement have perfectly known lengths. Minimum variance estimation for the unit vector attitude Kalman filter is studied in this work. An optimal multiplicative residual approach is presented. The proposed approach is compared with the classic additive residual attitude Kalman filter.

INTRODUCTION

The Kalman filter[1,2] is a widely used algorithm in spacecraft navigation. While the Kalman filter is usually employed to estimate vector quantities such as position or velocity, modifications to the classic algorithm exist to estimate attitude. One favorite spacecraft attitude representation is the quaternion-of-rotation.[3,4] Two common approaches to enforce the unit-norm constraint of the quaternion-of-rotation in the Kalman filter are the multiplicative extended Kalman filter (MEKF)[5] and the additive extended Kalman filter (AEKF).[6] Projection techniques and constrained Kalman filtering to enforce the quaternion normalization also exist.[7,8]

Direction measurements from attitude sensors are often provided as bearing angles. A unit vector can be created from the angles. While the Kalman filter can easily process the angular measurements, processing unit vectors is a widely adopted technique.[6,9] The QUEST measurement model[9] is a unit vector measurement model. More recently Cheng *et al.*[10] introduced a new measurement model. Both these models are additive; Mortari and Majji[11] introduced a multiplicative measurement model.

The covariance matrix of the additive measurement models is obtained through linearization assuming the measurement errors are small. This assumption is equivalent to making the measurement error perpendicular to the measurement itself. Therefore, the radial component of the error is identically zero and as a result the error covariance matrix is singular. Cheng *et al.*[10] avoid the problem by using the measurement error pseudoinverse in the

*Senior Member of the Technical Staff, GN&C Manned Space Systems, The Charles Stark Draper Laboratory, 17629 El Camino Real, Suite 470, Houston, Texas, 77058. E-mail: rzanetti@draper.com

information formulation of the Kalman filter. However, the information formulation leads to a minimum variance estimate only if both the estimation error covariance and the measurement error covariance are invertible. Catlin shows the general information estimate for singular measurement covariances [12, page 160] which differs from the minimum variance estimate. This work derives the general minimum variance estimate and shows that for this particular case the result is equivalent to the algorithm by Cheng *et al.*

The Kalman filter uses an additive residual (true measurement minus estimated measurement) to update the prior state once a measurement becomes available. Unit vector Kalman filters usually use the same technique producing good results. From a purely theoretical point of view however, subtracting two unit vectors to obtain a measurement residual is not satisfactory because the resulting vector has no physical meaning. Similarly to the AEKF, subtracting two unitary quantities provides an un-physical error representation, but the performance is not compromised by the operation.[13] While the MEKF has advantages and is widely used, it is philosophically unsatisfactory to reject the AEKF on the basis that the estimation error is un-physical and that the estimation error covariance is singular unless artificially increased in the radial direction, only to use a scheme that subtracts unit vector measurements and avoids singularity adding a radial component to the measurement covariance. In this work a multiplicative residual is employed in the MEKF, and it is shown that the additive residual is not necessary and redundant, a fact heretofore unknown.

MEASUREMENT MODEL

The unit vector measurement model used in this work is the one introduced by Cheng *et al.*[10] and is quickly presented here. The measured values are the apparent horizontal (a) and vertical (b) location of the observed direction in the camera sensitive element. Without loss of generality a unitary focal length is assumed. By defining the body-fixed camera frame having z along the boresight, x along the horizontal direction, and y along the vertical direction, the unit vector pointing to the apparent direction of the target in the body frame is given by

$$\mathbf{b} = \frac{1}{\sqrt{1+a^2+b^2}} \begin{bmatrix} -a \\ -b \\ 1 \end{bmatrix} . \tag{1}$$

The same vector coordinatized in the reference frame is denoted by $\mathbf{r}$. Denoting by δa and δb the measurement errors, the measured unit vector is given by

$$\tilde{\mathbf{b}} = \frac{1}{\sqrt{1+(a+\delta a)^2+(b+\delta b)^2}} \begin{bmatrix} -a-\delta a \\ -b-\delta b \\ 1 \end{bmatrix} \simeq \mathbf{b} + \mathbf{J} \begin{bmatrix} \delta a \\ \delta b \end{bmatrix} . \tag{2}$$

where $\mathbf{J}$ is the Jacobian of $\mathbf{b}$ and is given by[10]

$$\mathbf{J} = \frac{1}{\sqrt{1+a^2+b^2}} \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} - \frac{1}{1+a^2+b^2} \mathbf{b} \begin{bmatrix} a & b \end{bmatrix} . \tag{3}$$

For the purpose of this work, it is more convenient to write $\mathbf{J}$ in a different form. Define

$$\mathbf{I}_{3\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \tag{4}$$

then

$$\mathbf{J} = \frac{1}{\sqrt{1+a^2+b^2}} \left\{ -\mathbf{I}_{3\times 3} + \mathbf{b}\mathbf{b}^{\mathrm{T}} \right\} \mathbf{I}_{3\times 2} = \frac{1}{\sqrt{1+a^2+b^2}} [\mathbf{b}\times]^2 \mathbf{I}_{3\times 2}. \tag{5}$$

The covariance of the error in the focal plane is given by[10]

$$\mathrm{E}\left\{ \begin{bmatrix} \delta a \\ \delta b \end{bmatrix} \begin{bmatrix} \delta a \\ \delta b \end{bmatrix}^{\mathrm{T}} \right\} \triangleq \mathbf{R}^{FOCAL} = \frac{\sigma^2}{1+d(a^2+b^2)} \begin{bmatrix} (1+da^2)^2 & (dab)^2 \\ (dab)^2 & (1+db^2)^2 \end{bmatrix}, \tag{6}$$

where d is on the order of 1 and σ is known. The new measurement model introduced by Cheng *et al.* is given by

$$\mathbf{R}^{NEW} = \mathbf{J}\mathbf{R}^{FOCAL}\mathbf{J}^{\mathrm{T}}. \tag{7}$$

If a measurement occurs exactly at the boresight $a = b = 0$ and $\mathbf{b} = [0\ 0\ 1]^{\mathrm{T}}$. In this situation $\mathbf{R}^{FOCAL} = \sigma^2 \mathbf{I}_{2\times 2}$ and

$$\mathbf{R}^{NEW} = -\sigma^2 [\mathbf{b}\times]^2 = \sigma^2 (\mathbf{I}_{3\times 3} - \mathbf{b}\mathbf{b}^{\mathrm{T}}), \tag{8}$$

which is the QUEST measurement model.[9] The QUEST measurement model is valid for small field-of-view sensors for which is approximately equivalent to the model by Cheng *et al.* All the results presented in the following sections extend to the QUEST measurement model and the filter that employs it.[9]

MULTIPLICATIVE RESIDUAL APPROACH

Define $\bar{\mathbf{q}}_i^b$ as the true quaternion expressing the rotation from the reference frame i to a body-fixed frame b, and define $\hat{\bar{\mathbf{q}}}_i^b$ its estimate. Also let the three- dimensional estimation error[14] be $\delta\boldsymbol{\theta}$ such that

$$\bar{\mathbf{q}}_i^b = \bar{\mathbf{q}}(\delta\boldsymbol{\theta}) \otimes \hat{\bar{\mathbf{q}}}_i^b, \tag{9}$$

where $\bar{\mathbf{q}}(\delta\boldsymbol{\theta})$ is the quaternion representation of the rotation $\delta\boldsymbol{\theta}$. Define the following update

$$\delta\hat{\boldsymbol{\theta}}_k^+ = \delta\hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k(\tilde{\mathbf{b}}_k \times \hat{\mathbf{b}}_k), \tag{10}$$

where "$\times$" represents the vector cross product and "hat" denotes an estimated quantity. Using the multiplicative residual $\boldsymbol{\epsilon}_k = \tilde{\mathbf{b}}_k \times \hat{\mathbf{b}}_k$ is not completely new. It is used by Crassidis *et al.* in their trade study of contingency attitude determination algorithms.[15] The filter by Crassidis *et al.* differs from the proposed methodology because in their approach the rank-two covariance matrices are approximated by full-rank matrices, leading to a different update. The multiplicative residual is also used by Akella *et al.* for a continuous time nonlinear observer,[16] and is also derived by Reynolds as the extended Kalman filter mechanization of his attitude filter.[17]

A linear combination of the residual and the prior estimate (which is always equal to zero in the MEKF) is chosen as in the Kalman filter. The bigger the residual, the bigger the update. Zero residual means that the estimated measurement coincides with the actual measurement, therefore no update should occur. From the above discussion, it can be concluded that the multiplicative residual makes sense from a theoretically point of view because it is zero when the two vectors coincide, and it increases (in magnitude) as the sine of the angle between the two vectors. Therefore this filter would not work properly when the angle between the two vectors approaches or is greater than 90 degrees. Such large deviations, however, are beyond the scope of an EKF-like algorithm, which relies on linearization.

The residual can be rewritten as

$$\boldsymbol{\epsilon}_k = -[\hat{\mathbf{b}}_k \times] \left(\mathbf{T}(\delta\boldsymbol{\theta}^-)\hat{\mathbf{b}}_k + \frac{1}{\sqrt{1+a^2+b^2}}[\mathbf{b}\times]^2 \mathbf{I}_{3\times 2} \begin{bmatrix} \delta a \\ \delta b \end{bmatrix} \right) \tag{11}$$

$$\simeq -[\hat{\mathbf{b}}_k \times]^2 \delta\boldsymbol{\theta}_k^- - [\hat{\mathbf{b}}_k \times]\mathbf{J} \begin{bmatrix} \delta a \\ \delta b \end{bmatrix}. \tag{12}$$

where $\delta\boldsymbol{\theta}_k^-$ is the true prior attitude estimation error. Define

$$\mathbf{H}_k \triangleq -[\hat{\mathbf{b}}_k \times]^2 = \mathbf{I}_{3\times 3} - \hat{\mathbf{b}}_k \hat{\mathbf{b}}_k^{\mathrm{T}} = \mathbf{H}_k^{\mathrm{T}}. \tag{13}$$

The true posterior error is

$$\delta\boldsymbol{\theta}_k^+ = (\mathbf{I}_{3\times 3} - \mathbf{K}_k \mathbf{H}_k)\delta\boldsymbol{\theta}_k^- - \mathbf{K}_k [\hat{\mathbf{b}}_k \times] \mathbf{J}_k \begin{bmatrix} \delta a \\ \delta b \end{bmatrix}. \tag{14}$$

Define the estimation error covariance $\mathbf{P}$ as

$$\mathbf{P}_k = \mathrm{E}\left\{ \delta\boldsymbol{\theta}_k \delta\boldsymbol{\theta}_k^{\mathrm{T}} \right\}. \tag{15}$$

Assuming the measurement error is white and uncorrelated from other error sources the posterior covariance is given by

$$\mathbf{P}_k^+ = (\mathbf{I}_{3\times 3} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_k^- (\mathbf{I}_{3\times 3} - \mathbf{K}_k \mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}}. \tag{16}$$

The goal is to choose the gain $\mathbf{K}_k$ that minimizes the trace of the posterior covariance ($\mathrm{trace}\,\mathbf{P}_k^+$) providing the minimum variance estimate. The optimal gain is given by[12]

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}})^{\dagger}, \tag{17}$$

where the symbol "†" indicates the Moore-Penrose pseudoinverse. It does not matter that the covariance $\mathbf{W}_k$ of the residual is singular. The optimal gain can be found using the pseudoinverse instead of the regular inverse.

$$\mathbf{W}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}}. \tag{18}$$

A small complication arises from the fact that $\mathbf{W}_k$ is neither full row nor full column, therefore it might seem that singular value decomposition is needed to calculate the pseudoinverse, which would reduce the practical usefulness of the algorithm. However, the algorithm proposed here calculates the optimal gain without singular value decomposition.

Lemma 1. *Define the unit vectors* $\mathbf{n}_1$ *and* $\mathbf{n}_2$ *such that* $\{\hat{\mathbf{b}}_k, \mathbf{n}_1, \mathbf{n}_2\}$ *is an orthonormal triad. Define also*

$$\mathbf{S} = \begin{bmatrix} \mathbf{n}_1 & \mathbf{n}_2 \end{bmatrix} . \tag{19}$$

Then

$$\mathbf{H}_k = \mathbf{S}\mathbf{S}^{\mathrm{T}}. \tag{20}$$

Proof. To show that the two matrices are the same it is sufficient to show that a non singular matrix $\mathbf{A}$ exists such that $\mathbf{H}_k\mathbf{A} = \mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{A}$. An appropriate choice is

$$\mathbf{A} = \begin{bmatrix} \hat{\mathbf{b}}_k & \mathbf{n}_1 & \mathbf{n}_2 \end{bmatrix} , \tag{21}$$

which is clearly of full rank since $\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}$.

$$\mathbf{H}_k\mathbf{A} = -\begin{bmatrix} [\hat{\mathbf{b}}_k\times]^2\hat{\mathbf{b}}_k & [\hat{\mathbf{b}}_k\times]^2\mathbf{n}_1 & [\hat{\mathbf{b}}_k\times]^2\mathbf{n}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{n}_1 & \mathbf{n}_2 \end{bmatrix} . \tag{22}$$

$$\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{A} = \mathbf{S}\begin{bmatrix} \mathbf{n}_1^{\mathrm{T}}\hat{\mathbf{b}}_k & \mathbf{n}_1^{\mathrm{T}}\mathbf{n}_1 & \mathbf{n}_1^{\mathrm{T}}\mathbf{n}_2 \\ \mathbf{n}_2^{\mathrm{T}}\hat{\mathbf{b}}_k & \mathbf{n}_2^{\mathrm{T}}\mathbf{n}_1 & \mathbf{n}_2^{\mathrm{T}}\mathbf{n}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{n}_1 & \mathbf{n}_2 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{23}$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{n}_1 & \mathbf{n}_2 \end{bmatrix} . \tag{24}$$

□

Property 1. If matrix $\mathbf{A}$ is of full column rank, and $\mathbf{B}$ is of full row rank, then

$$(\mathbf{A}\mathbf{B})^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger}. \tag{25}$$

Lemma 2. *Using the definitions of Lemma 1 then*

$$\mathbf{S}^{\dagger} = \mathbf{S}^{\mathrm{T}}. \tag{26}$$

Proof. The proof comes directly from the definition of pseudoinverse which must satisfy

1. $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}$
2. $\mathbf{A}^{\dagger}\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{A}^{\dagger}$
3. $(\mathbf{A}\mathbf{A}^{\dagger})^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\dagger}$
4. $(\mathbf{A}^{\dagger}\mathbf{A})^{\mathrm{T}} = \mathbf{A}^{\dagger}\mathbf{A}$

□

Theorem 1. *The optimal gain of Eq. (17) is given by*

$$\mathbf{K}_k = \mathbf{P}_k^{-}\mathbf{S}\{\mathbf{S}^{\mathrm{T}}(\mathbf{P}_k^{-} + [\hat{\mathbf{b}}_k\times]\mathbf{R}_k^{NEW}[\hat{\mathbf{b}}_k\times]^{\mathrm{T}})\mathbf{S}\}^{-1}\mathbf{S}^{\mathrm{T}} \tag{27}$$

Proof.

$$
\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}})^{\dagger} && (28)\\
&= \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \{\mathbf{S}\mathbf{S}^{\mathrm{T}} (\mathbf{P}_k^- + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}}) \mathbf{S}\mathbf{S}^{\mathrm{T}}\}^{\dagger} \quad \text{from Lemma 1} && (29)\\
&= \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} (\mathbf{S}^{\mathrm{T}})^{\dagger} \{\mathbf{S}^{\mathrm{T}} (\mathbf{P}_k^- + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}}) \mathbf{S}\}^{\dagger} \, \mathbf{S}^{\dagger} \quad \text{from Property 1} && (30)\\
&= \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{S} \{\mathbf{S}^{\mathrm{T}} (\mathbf{P}_k^- + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}}) \mathbf{S}\}^{-1} \mathbf{S}^{\mathrm{T}} \quad \text{from Lemma 2} && (31)\\
&= \mathbf{P}_k^- \mathbf{S} \{\mathbf{S}^{\mathrm{T}} (\mathbf{P}_k^- + [\hat{\mathbf{b}}_k \times] \mathbf{R}_k^{NEW} [\hat{\mathbf{b}}_k \times]^{\mathrm{T}}) \mathbf{S}\}^{-1} \mathbf{S}^{\mathrm{T}}. && (32)
\end{aligned}
$$

The fact that the pseudoinverse coincides with the inverse when the latter exists is also used. Finally the last equality holds because

$$
\mathbf{H}_k^{\mathrm{T}} \mathbf{S} = -\begin{bmatrix} [\hat{\mathbf{b}}_k \times]^2 \mathbf{n}_1 & [\hat{\mathbf{b}}_k \times]^2 \mathbf{n}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{n}_1 & \mathbf{n}_2 \end{bmatrix} = \mathbf{S}. \tag{33}
$$

□

Notice that matrix $\mathbf{S}^{\mathrm{T}}[\hat{\mathbf{b}}_k \times]\mathbf{R}_k^{NEW}[\hat{\mathbf{b}}_k \times]^{\mathrm{T}}\mathbf{S} \in \Re^{2\times 2}$ is of full rank as long as $\mathbf{b}$ is not perpendicular to the boresight. Clearly no sensor can have a 180 degree field-of-view because it would require an infinitely large sensitive element (given any finite focal length).

The covariance update is given by the Joseph formula of Eq. (16) and using Eq. (33) to notice that $\mathbf{K}_k \mathbf{H}_k = \mathbf{K}_k$.

$$
\mathbf{P}^+ = (\mathbf{I}_{3\times 3} - \mathbf{K}_k)\mathbf{P}^-(\mathbf{I}_{3\times 3} - \mathbf{K}_k)^{\mathrm{T}} + \mathbf{K}[\hat{\mathbf{b}}_k \times]\mathbf{R}_k^{NEW}[\hat{\mathbf{b}}_k \times]^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}. \tag{34}
$$

COMPARISON WITH LARGE FIELD-OF-VIEW MODEL

The filter proposed in Ref. 10 is given by

$$
\begin{aligned}
\tilde{\mathbf{H}}_k &= [\hat{\mathbf{b}}_k \times] && (35)\\
(\mathbf{P}^+)^{-1} &= (\mathbf{P}^+)^{-1} + \tilde{\mathbf{H}}_k^{\mathrm{T}} (\mathbf{R}_k^{NEW})^{\dagger} \tilde{\mathbf{H}}_k && (36)\\
\tilde{\mathbf{K}}_k &= \mathbf{P}_k^+ \tilde{\mathbf{H}}_k^{\mathrm{T}} (\mathbf{R}_k^{NEW})^{\dagger} && (37)\\
\delta\hat{\boldsymbol{\theta}}_k^+ &= \delta\hat{\boldsymbol{\theta}}_k^- + \tilde{\mathbf{K}}_k (\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k). && (38)
\end{aligned}
$$

The linear unbiased minimum variance estimator is given by the Kalman filter. The minimum variance estimator is equivalent to the information formulation when the hypotheses of the matrix inversion lemma are satisfied. These hypotheses are that both the estimation error covariance and the measurement error covariance are non singular. When the hypotheses are not satisfied, the information formulation of the Kalman filter does not generally provide the minimum variance estimate.[12] Therefore the procedure followed in Ref. 10 does not guarantee that their estimate is optimal in a minimum variance sense. The actual minimum variance estimate is given by

$$
\begin{aligned}
\tilde{\mathbf{W}}_k &= \tilde{\mathbf{H}}_k \mathbf{P}_k^- \tilde{\mathbf{H}}_k^{\mathrm{T}} + \mathbf{R}_k^{NEW} && (39)\\
\tilde{\mathbf{K}}_k &= \mathbf{P}_k^- \tilde{\mathbf{H}}_k^{\mathrm{T}} \tilde{\mathbf{W}}_k^{\dagger} && (40)\\
\delta\hat{\boldsymbol{\theta}}_k^+ &= \delta\hat{\boldsymbol{\theta}}_k^- + \tilde{\mathbf{K}}_k (\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k) && (41)\\
\mathbf{P}^+ &= (\mathbf{I}_{3\times 3} - \tilde{\mathbf{K}}_k \tilde{\mathbf{H}}_k)\mathbf{P}^-(\mathbf{I}_{3\times 3} - \tilde{\mathbf{K}}_k \tilde{\mathbf{H}}_k)^{\mathrm{T}} + \tilde{\mathbf{K}}_k \mathbf{R}^{NEW} \tilde{\mathbf{K}}_k^{\mathrm{T}}. && (42)
\end{aligned}
$$

From Ref. [10, Eq. (27)]:

$$(\tilde{\mathbf{W}}_k + c\hat{\mathbf{b}}_k\hat{\mathbf{b}}_k^{\mathrm{T}})^{-1} = \tilde{\mathbf{W}}_k^{\dagger} + \frac{1}{c}\hat{\mathbf{b}}_k\hat{\mathbf{b}}_k^{\mathrm{T}}, \tag{43}$$

and noticing that $\tilde{\mathbf{H}}_k^{\mathrm{T}}\hat{\mathbf{b}} = \mathbf{0}$, it follows that

$$\tilde{\mathbf{K}}_k = \mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}}\tilde{\mathbf{W}}_k^{\dagger} = \mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}}(\tilde{\mathbf{W}}_k + c\hat{\mathbf{b}}_k\hat{\mathbf{b}}_k^{\mathrm{T}})^{-1}. \tag{44}$$

By adding a non-zero component to the measurement error covariance along the radial direction $\mathcal{R}_k^{NEW} = \mathbf{R}_k^{NEW} + c\hat{\mathbf{b}}_k\hat{\mathbf{b}}_k^{\mathrm{T}}$, the new residual covariance becomes

$$\tilde{\mathbf{W}}_k + c\hat{\mathbf{b}}_k\hat{\mathbf{b}}_k^{\mathrm{T}} = \tilde{\mathbf{H}}_k\mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}} + \mathcal{R}_k^{NEW}, \tag{45}$$

which satisfies the hypotheses of the matrix inversion lemma because both $\mathbf{P}_k^{-}$ and $\mathcal{R}_k^{NEW}$ are of full rank. Therefore the information formulation can be used and the equations by Cheng *et al.* follow. Simply replacing the inverse with the pseudoinverse in the information formulation of the Kalman filter is not sufficient to prove optimality. The derivation presented here proves that the method by Cheng *et al.* is optimal in a minimum variance sense. The method is also equivalent to the one proposed in the previous section as it will now be shown.

The gain by Cheng *et al.* is given by

$$\tilde{\mathbf{K}}_k = \mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}}(\tilde{\mathbf{H}}_k\mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}} + \mathbf{R}_k^{NEW})^{\dagger} = -\mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}}[\hat{\mathbf{b}}\times]^2(\tilde{\mathbf{H}}_k\mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}} + \mathbf{R}_k^{NEW})^{\dagger} \tag{46}$$

since $\tilde{\mathbf{H}}_k^{\mathrm{T}} = -[\hat{\mathbf{b}}_k\times] = -\tilde{\mathbf{H}}_k^{\mathrm{T}}[\hat{\mathbf{b}}_k\times]^2$. Multiplying both sides by $[\hat{\mathbf{b}}_k\times]^{\mathrm{T}}$

$$\tilde{\mathbf{K}}_k[\hat{\mathbf{b}}_k\times]^{\mathrm{T}} = -\mathbf{P}_k^{-}\mathbf{H}_k^{\mathrm{T}}[\hat{\mathbf{b}}_k\times](\tilde{\mathbf{H}}_k\mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}} + \mathbf{R}_k^{NEW})^{\dagger}[\hat{\mathbf{b}}_k\times]^{\mathrm{T}}, \tag{47}$$

since $\tilde{\mathbf{H}}_k^{\mathrm{T}}[\hat{\mathbf{b}}_k\times]^2 = -[\hat{\mathbf{b}}_k\times]^3 = \mathbf{H}_k^{\mathrm{T}}[\hat{\mathbf{b}}_k\times]$. Using

$$[\hat{\mathbf{b}}_k\times]^{\dagger} = [\hat{\mathbf{b}}_k\times]^{\mathrm{T}}, \tag{48}$$

it follows that

$$\tilde{\mathbf{K}}_k[\hat{\mathbf{b}}\times]^{\mathrm{T}} = -\tilde{\mathbf{K}}_k\tilde{\mathbf{H}}_k = -\mathbf{P}_k^{-}\mathbf{H}_k^{\mathrm{T}}([\hat{\mathbf{b}}\times]\tilde{\mathbf{H}}_k\mathbf{P}_k^{-}\tilde{\mathbf{H}}_k^{\mathrm{T}}[\hat{\mathbf{b}}\times]^{\mathrm{T}} + [\hat{\mathbf{b}}\times]\mathbf{R}_k^{NEW}[\hat{\mathbf{b}}\times]^{\mathrm{T}})^{\dagger} = -\mathbf{K}_k. \tag{49}$$

The covariance update is given by Eq. (16)

$$\mathbf{P}^{+} = (\mathbf{I}_{3\times3} - \mathbf{K}_k)\mathbf{P}^{-}(\mathbf{I}_{3\times3} - \mathbf{K}_k)^{\mathrm{T}} + \mathbf{K}_k[\hat{\mathbf{b}}_k\times]\mathbf{R}_k^{NEW}[\hat{\mathbf{b}}_k\times]^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} \tag{50}$$

$$= (\mathbf{I}_{3\times3} - \tilde{\mathbf{K}}_k\tilde{\mathbf{H}}_k)\mathbf{P}^{-}(\mathbf{I}_{3\times3} - \tilde{\mathbf{K}}_k\tilde{\mathbf{H}}_k)^{\mathrm{T}} + \tilde{\mathbf{K}}_k\mathbf{R}^{NEW}\tilde{\mathbf{K}}_k^{\mathrm{T}}, \tag{51}$$

therefore the two methods lead to the same covariance update.

The multiplicative update is given by

$$\delta\hat{\boldsymbol{\theta}}_k^{+} = \delta\hat{\boldsymbol{\theta}}_k^{-} - \mathbf{K}_k[\hat{\mathbf{b}}_k\times]\tilde{\mathbf{b}}_k = \delta\hat{\boldsymbol{\theta}}_k^{-} - \mathbf{K}_k[\hat{\mathbf{b}}_k\times](\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k) \tag{52}$$

$$= \delta\hat{\boldsymbol{\theta}}_k^{-} - \tilde{\mathbf{K}}_k[\hat{\mathbf{b}}_k\times]^2(\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k) = \delta\hat{\boldsymbol{\theta}}_k^{-} + \tilde{\mathbf{K}}_k(\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k), \tag{53}$$

therefore the two methods lead to the same state update. The last equality of Eq. (53) holds because $[\mathbf{b}_k\times] = -[\mathbf{b}_k\times]^3$ and

$$\tilde{\mathbf{K}}_k = \mathbf{P}_k^+\tilde{\mathbf{H}}_k^{\mathrm{T}}(\mathbf{R}_k^{NEW})^\dagger = \mathbf{P}_k^+\tilde{\mathbf{H}}_k^{\mathrm{T}}(\mathbf{J}\mathbf{R}_k^{FOCAL}\mathbf{J}^{\mathrm{T}})^\dagger \tag{54}$$

$$= \mathbf{P}_k^+\tilde{\mathbf{H}}_k^{\mathrm{T}}(\mathbf{J}^{\mathrm{T}})^\dagger(\mathbf{R}_k^{FOCAL})^{-1}(\mathbf{J})^\dagger, \tag{55}$$

where $\mathbf{J}$ is evaluated at the estimated measurement. But from Eqs. (48) and (5)

$$\mathbf{J}^\dagger = \sqrt{1+a^2+b^2}(\mathbf{I}_{3\times 2})^\dagger([\mathbf{b}\times]^2)^\dagger = \sqrt{1+a^2+b^2}(\mathbf{I}_{2\times 3})[\mathbf{b}\times]^2. \tag{56}$$

Therefore $\tilde{\mathbf{K}}_k = \mathbf{C}_k[\hat{\mathbf{b}}_k\times]$, where $\mathbf{C}_k$ is some matrix. This last identity also implies that the additive update

$$\delta\hat{\boldsymbol{\theta}}_k^+ = \tilde{\mathbf{K}}_k(\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k) \tag{57}$$

is unnecessary and can be reduced to

$$\delta\hat{\boldsymbol{\theta}}_k^+ = \tilde{\mathbf{K}}_k\tilde{\mathbf{b}}_k \tag{58}$$

where the fact that $\delta\hat{\boldsymbol{\theta}}_k^-$ is identically zero is also used. The equivalence shown in this section expands to the filter based on the QUEST measurement model proposed in Ref. 9, since that filter is equivalent to the one by Cheng *et al.* as the field-of-view approaches zero. The filter in Ref. 9 adds artificial noise along the radial direction to avoid matrix singularity issues.

CONCLUSION

A unit-vector quaternion Kalman filter with multiplicative residual is derived. It is shown that the new algorithm is equivalent to an existing scheme employing additive residuals. The algorithms are shown to be optimal in a minimum variance sense. It is also shown that the additive residual is redundant. The proposed algorithm requires the inversion of a two-by-two matrix while the original algorithm requires the inversion of an n-by-n matrix, where n is the size of the state vector. The algorithms are equivalent and can be used interchangeably, however, the new multiplicative residual scheme solves an existing incongruence of multiplicative extended Kalman filters (MEKF) using unit-vector measurements. While the additive extended Kalman filter (AEKF) is shown to perform satisfactorily in practical situations, some objections are raised to prefer the use of the MEKF instead. One such objection is that the AEKF inevitably possesses a singular covariance that is avoided by adding artificial noise to the quaternion magnitude direction. Another objection is that subtracting unit quaternions to form a quaternion error is un-physical. While posing these objections, MEKF designs employing unit-vector measurements often use very similar techniques. The state update of the MEKF is obtained using an additive residual measurement that subtracts two unit vectors. The additive residual is introduced by Kalman in the contest of vector spaces. Clearly subtracting two unit-vector measurements produces an un-physical residual. The residual's covariance of the unit-vector MEKF is also singular, since the length of the measurement is perfectly known. This problem is usually overcome by adding artificial error along the radial direction. The author believes the above objections are not as important as designing a filter that practically produces good results, as both the AEKF and the unit-vector MEKF do. This work, however, is aimed to reconcile the mentioned shortcomings with the existing design of the unit-vector MEKF.

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, March 1960, pp. 35–45.

[2] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction," *Journal of Basic Engineering*, Vol. 83, March 1961, pp. 95–108.

[3] J. B. Kuipers, *Quaternions and Rotation Sequences*. Princeton, NJ: Princeton University Press, 1999.

[4] F. L. Markley, "Attitude Estimation or Quaternion Estimation?," *The Journal of the Astronautical Sciences*, Vol. 52, January–June 2004, pp. 221–238.

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

[6] I. Y. Bar-Itzhack and Y. Oshman, "Attitude Determination from Vector Observations: Quaternion Estimation," *IEEE Transaction on Aerospace and Electronic Systems*, Vol. 21, January 1985, pp. 128–135.

[7] D. Simon and T. L. Chia, "Kalman Filtering with State Equality Constraints," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 38, January 2002, pp. 128–136.

[8] R. Zanetti, M. Majji, R. H. Bishop, and D. Mortari, "Norm-Constrained Kalman Filtering," *Journal of Guidance, Control, and Dynamics*, Vol. 32, September–October 2009.

[9] M. D. Shuster, "Kalman Filtering of Spacecraft Attitude and the QUEST Model," *The Journal of the Astronautical Sciences*, Vol. 38, July–September 1990, pp. 377–393.

[10] Y. Cheng, J. L. Crassidis, and F. L. Markley, "Attitude Estimation for Large Field-of-View Sensors," *The Journal of the Astronautical Sciences*, Vol. 54, July–December 2006, pp. 433–448.

[11] D. Mortari and M. Majji, "Multiplicative Measurement Model and Single-Point Attitude Estimation," *Proceedings of the AAS F. Landis Markley Astronautics Symposium*, Cambridge, MD, June 29–July 2 2008, pp. 51–69. AAS 08-263.

[12] D. E. Catlin, *Estimation, Control, and the Discrete Kalman Filter*. No. 71 in Applied Mathematical Sciences, New York: Springer-Verlag, 1989.

[13] J. Deutschmann, I. Bar-Itzhack, and K. Galal, "Quaternion Normalization in Spacecraft Attitude Determination," *AIAA/AAS Astrodynamics Specialist Conference*, August 1992.

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

[15] J. L. Crassidis, S. F. Andrews, F. L. Markley, and K. Ha, "Contingency Designs for Attitude Determination of TRMM," *NASA/GSFC Flight Mechanics Symposium*, NASA/CP 3299, Greenbelt, MD, May 1995.

[16] M. R. Akella, D. Seo, and R. Zanetti, "Attracting Manifolds for Attitude Estimation in Flatland and Otherlands," *The Journal of the Astronautical Sciences*, Vol. 54, July–December 2006, pp. 635–655.

[17] R. G. Reynolds, "Asymptotically Optimal Attitude Filtering with Guaranteed Convergence," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 1, 2008, pp. 114–122.