Journal of Guidance, Control, and Dynamics
Vol. 31, No. 6, November–December 2008

Asymptotic Behavior of the Estimation Error Covariance of Quaternion Estimators

Avishy Carmi* and Yaakov Oshman†
Technion—Israel Institute of Technology, Haifa 32000, Israel

DOI: 10.2514/1.35781

The asymptotic behavior of the estimation error covariance of quaternion estimators is mathematically examined. It is proved that the condition number of the asymptotic covariance matrix is of the order of the inverse of its largest eigenvalue, so that this matrix becomes asymptotically ill-conditioned as its trace tends to zero. Nevertheless, it is proved that the aforementioned asymptotic behavior cannot be captured by low-order Taylor approximations of the covariance, such as the one computed by the extended Kalman filter. Geometrical interpretation of the results is provided, using tools from differential geometry. The analytical results are demonstrated via a simulation study using the recently introduced quaternion particle filter and the additive quaternion extended Kalman filter.

I. Introduction

Being used in many attitude estimation algorithms, the rotation quaternion is perhaps the most common spacecraft attitude representation. The most attractive feature of this attitude specifier is that it is not singular for any rotation. Moreover, its kinematic equation is linear and the computation of the associated attitude matrix involves only algebraic expressions. However, the quaternion representation is not minimal because it is four-dimensional. This leads to a normalization constraint that has to be addressed in filtering algorithms.

In the early 1980s researchers began using the celebrated extended Kalman filter (EKF) for spacecraft attitude filtering algorithms. The EKF (and its variants) has become the most widely applied attitude estimation algorithm ever since. The EKF is an extension of the Kalman filter (KF) for nonlinear systems. Like the KF, the EKF represents the filtering probability distribution function (pdf) via its first two moments (namely, mean and covariance) only. Whereas the KF yields the (optimal) minimum mean square error (MMSE) estimate in the case of linear Gaussian models, the EKF yields suboptimal (in the MMSE sense) estimates, at best, when applied in nonlinear/non-Gaussian systems.

Employing the EKF or a similar filtering mechanization that computes the second-order statistics of the filtering pdf using a quaternion attitude representation has lead to a debate among the attitude determination community. At the core of the debate is the question of whether the quaternion estimation error covariance matrix is singular or ill-conditioned due to the quaternion's unit-norm constraint. A notable example is [1], that assumes that the 4×4 quaternion estimation error covariance matrix must be singular and surveys reduced-order algorithms that maintain this singularity. The suggested reduced-order algorithms are based on three approaches: the first approach uses the transition matrix of the state error vector to obtain a reduced-order error covariance representation. The second approach consists of omitting one of the quaternion components to obtain a truncated covariance expression. Finally, the third approach uses a compositional incremental quaternion error, which leads to a covariance representation similar to that of the first approach.

The quaternion estimation error covariance singularity is mentioned in [2,3] as a direct consequence of the quaternion norm constraint. Although the algorithm suggested in [2] is not a KF variant, the singularity issue is pointed out as validating the use of the reduced-order EKFs of [1] and the minimum model error estimator of [2]. In [4] a nonlinear predictive filtering approach is presented for gyroless spacecraft attitude estimation. One of the advantages of the predictive filter over the EKF is its ability to maintain the quaternion norm constraint. Reference [4] maintains that the EKF quaternion estimation error covariance is strictly singular due to the norm constraint.

Reference [5] presents a quaternion-based autonomous attitude estimation and control architecture for the Ørsted satellite. Addressing the EKF's estimation error covariance singularity, the attitude determination algorithm is chosen to be a reduced-order EKF, using the incremental quaternion error approach of [1].

In a recent work [6], an unscented KF (UKF) has been proposed for the estimation of the rotation quaternion. The UKF does possess a reported advantage over the EKF with regard to dealing with strongly nonlinear systems, because it avoids the linearization process associated with the EKF. However, because using the UKF directly with the quaternion attitude parameterization would also yield a nonunit norm quaternion estimate, the authors of [6] chose to work with a generalized three-dimensional attitude representation, still using the quaternion for updates to maintain the normalization constraint. Further justifying the use of a three-component attitude representation, [6] states that the quaternion estimation error covariance is ill-conditioned and may also be singular when a linear computation, such as the EKF, is involved.

An extensive survey of EKF-based quaternion filtering strategies is presented in [7]. Divided into two broad classes, the filtering algorithms representing the attitude errors using minimal-parameter representations are compared with the one that directly estimates the four-component quaternion. As part of its conclusions, [7] asserts that the latter algorithm suffers from numerical instability originating in an asymptotically singular covariance matrix.

Reference [8] suggests a unique EKF-based algorithm for spinning spacecraft attitude estimation. Resolving the angular-momentum vector in both the inertial-reference and body-frame coordinate systems while using a three-axis magnetometer as its primary sensing device, the derived filter is able to compute attitude estimates. As part of its mechanization, the proposed algorithm addresses the problem of enforcing both angular-momentum representations to have the same norm. Consequently, [8] resorts to a reduced-order covariance representation for maintaining its singularity, thereby corroborating with [1].

Presented as Paper 6814 at the AIAA Guidance, Navigation, and Control Conference and Exhibit, Hilton Head, SC, 20–23 August 2007; received 21 November 2007; revision received 27 April 2008; accepted for publication 28 April 2008.

*Doctoral Student, Department of Aerospace Engineering; avishy@aerodyne.technion.ac.il. Member AIAA.

†Professor, Department of Aerospace Engineering; Holder, Louis and Helen Rogow Chair in Aeronautical Engineering; yaakov.oshman@technion.ac.il. Fellow AIAA.

The quaternion covariance singularity/ill-conditioning is circumvented in [9,10]. The former paper suggests a new method for deriving quaternion estimates on the unit 3-sphere. A major part of this work is devoted to a new quaternion probabilistic model expressed solely in terms of the quaternion second moment. A sampling procedure is then developed, adopting the proposed model for use by Monte Carlo type of algorithms. Reference [10] parameterizes the attitude using the direction cosine matrix (DCM) while deriving an analytic optimal filtering method. The time-propagation and measurement-update stages are implemented using the Fokker–Planck stochastic differential equation and Bayes formula, respectively. The DCM optimal estimates are then obtained based on either minimizing a cost function expressed exclusively using the attitude estimation error Frobenius norm on SO(3) or minimizing a negative-log-likelihood function related to the probability density on SO(3).

In other papers the covariance singularity issue is either regarded as a misunderstanding or is completely avoided. These papers mostly rely on the fact that practical implementation of the EKF using a quaternion attitude representation has never, in the past, yielded a singular estimation error covariance. As an example, the authors of [11] assume no such singularity while incorporating a normalization step into their quaternion-based EKF algorithm. Reference [12] illustrates the regularity of the EKF covariance matrix through several examples that consist of estimating random variables with a functional relationship. It then analyzes the quaternion estimation problem, reaching the conclusion that the discrepancy stems from the fact that a functional relationship does exist between the quaternion components, but not between the components of its incremental estimation error.

In a related work [13], a method is presented for enforcing an algebraic constraint in a quaternion-based EKF. Similar to [11], the EKF formulation suggested by [13] involves the full quaternion vector as part of its state. A correction step is then devised for the EKF measurement-update stage, thereby increasing its stability. Finally, the efficiency of the algorithm is demonstrated via a numerical simulation study. No covariance ill-conditioning or singularity has been detected in this study.

As part of an extensive comparison between the two common approaches for EKF-based quaternion estimation of [1,11], [14] states the conditions under which the quaternion estimation error covariance is expected to become nearly singular and claims that the EKF-computed covariance will not become ill-conditioned in the presence of process noise. However, in the absence of process noise, [14] claims that covariance ill-conditioning is an expected characteristic of the additive EKF (AEKF) algorithm of [11].

Recently, innovative methods have been proposed for spacecraft attitude estimation. In [15], a particle filter (PF) was implemented to sequentially estimate the attitude quaternion from vector observations. Also known as sequential Monte Carlo methods, PFs are algorithms implementing a recursive Bayesian model using simulation-based methods (see [16]). Avoiding the underlying assumptions of the KF (namely, that the state space is linear and Gaussian), these rather general and flexible methods enable solving for the entire posterior probability distributions of the unknown variables (on which all inference on these variables is based) within a Bayesian framework, exploiting the dramatic recent increase in computer power. Contrary to KF extensions, the solutions obtained using PF algorithms are numerical approximations to the optimal (in the Bayesian sense) solutions, which can be made arbitrarily close to the exact solutions by increasing the number of particles (samples) involved in the computation. It should be emphasized that implementing a PF for the attitude quaternion estimation completely avoids the covariance singularity issue, because the entire filtering distribution is represented via particles and not by its statistical moments. Thus, the error covariance matrix is not a formal part of the algorithm, although it can always be computed, if needed, using the filtering distribution (represented by its particles).

An extensive evaluation of the properties of the estimation error covariance matrix computed using the quaternion PF (QPF) of [15] has confirmed that it becomes ill-conditioned during the estimation procedure. This finding has provided the motivation for the work presented here.

This paper mathematically examines the asymptotic behavior of the quaternion estimation error covariance matrix. It proves that the condition number of the asymptotic covariance matrix is of the order of the inverse of its largest eigenvalue. Furthermore, it proves that the aforementioned asymptotic behavior cannot be captured by low-order Taylor approximations of this matrix. Conclusions are drawn for the EKF case. The essential conclusions are demonstrated by an extensive simulation study that uses both the QPF of [15] and the AEKF of [11].

To set the stage for some of the key issues tackled in this paper, consider the following simple example. Let $\boldsymbol{x}$, a random vector (rv) taking values on the unit circle, be defined as

$$\boldsymbol{x} = [\cos\theta, \sin\theta]^T, \qquad \theta \sim U(-\tfrac{1}{2}a, \tfrac{1}{2}a), \qquad a \in \mathbb{R} \tag{1}$$

where $U(c_1, c_2)$ denotes a uniform distribution over $[c_1, c_2]$. The eigenvalues of the covariance matrix $\operatorname{cov}(\boldsymbol{x}) = \operatorname{diag}\{\lambda_1, \lambda_2\}$ are

$$\lambda_1 = \frac{1}{2} + \frac{1}{2}\frac{\sin a}{a} - \left(\frac{\sin(a/2)}{a/2}\right)^2, \qquad \lambda_2 = \frac{1}{2} - \frac{1}{2}\frac{\sin a}{a} \tag{2}$$

Thus, although $\boldsymbol{x}$ is a constrained rv, its covariance is not, in general, singular or ill-conditioned (setting $a = 2\pi$ yields $\lambda_1 = \lambda_2 = 0.5$). However,

$$\lim_{a\to 0}\frac{\lambda_1}{\lambda_2} = 0 \tag{3}$$

implying that $\operatorname{cov}(\boldsymbol{x})$ does become asymptotically ill-conditioned as $a \to 0$.

The four-dimensional counterpart of the rv $\boldsymbol{x}$ is the rotation quaternion. It turns out that the covariance matrix associated with the quaternion estimation error possesses the asymptotic property underlined in this example. Furthermore, recalling that in a typical estimation procedure the trace of the covariance matrix decreases rapidly, this property indicates that ill-conditioning is inevitable in optimal quaternion estimation (however, it cannot be used to indicate such behavior in suboptimal estimators).

The remainder of this paper is organized as follows. The next section reviews some mathematical concepts from asymptotic analysis and examines the behavior of the second-moment matrix of a nonlinearly constrained random vector. The main results of the paper are presented in Sec. III, in which the conclusions from the analysis of Sec. II are used to infer the behavior of the true quaternion estimation error covariance and of the AEKF covariance. Section IV presents a geometric interpretation of the results of the previous section using differential geometry tools. Section V presents the results of a numerical simulation study that is carried out to validate the analytical conclusions obtained throughout this paper. Concluding remarks are offered in the last section. For improved readability, several auxiliary results and proofs are deferred to the Appendices.

II. Second Moment of a Constrained Random Vector

The asymptotic behavior of the quaternion error covariance can be regarded as part of the more general problem of evaluating the statistical properties of a constrained rv. Directed at the quaternion covariance problem, this section is concerned with the asymptotic behavior of the second-moment matrix as its trace tends to zero. For completeness, the analysis is preceded by some definitions.

A. Order Symbols in Asymptotic Analysis

The following definitions are taken from [17] (pp. 2–3).

Definition 1. Let X be a metric space and let $x_0 \in X$. A *pointed neighborhood* of x_0 is a set of the form $V \setminus \{x_0\}$, where V is a neighborhood of x_0.

Definition 2. Let $f(x)$ and $g(x)$ be functions defined in a pointed neighborhood V of x_0. Then $f(x)$ is $\mathcal{O}(g(x))$ as $x \to x_0$, denoted as

$$f(x) = \mathcal{O}(g(x)) \qquad \text{as } x \to x_0 \tag{4}$$

if there exists a pointed neighborhood V of x_0 and a constant $0 < c < \infty$ such that

$$|f(x)| \leq c|g(x)|, \qquad x \in V \tag{5}$$

If $g(x)$ does not vanish near x_0, then the relation (4) is equivalent to the condition

$$\limsup_{x \to x_0} \left| \frac{f(x)}{g(x)} \right| < \infty \tag{6}$$

Order symbols can be used to effectively convey the dominance of mathematical forms without writing them explicitly. Thus, the order of any mathematical entity can be expressed by manipulating the order of its elements. The basic properties of order symbols can be found in the literature on asymptotic analysis [17]. Some useful properties of the order symbol $\mathcal{O}(\cdot)$ that are extensively used in this work are reviewed below. The notation is the one commonly used in asymptotic analysis; the meaning of $\mathcal{O}(g)$ is that some function f is $\mathcal{O}(g)$. Therefore, $\mathcal{O}(g)\mathcal{O}(g) = \mathcal{O}(g^2)$ implies that if $f_1 = \mathcal{O}(g)$ and $f_2 = \mathcal{O}(g)$, then also $f_1 f_2 = \mathcal{O}(g^2)$.

1) $\mathcal{O}(g)\mathcal{O}(f) = \mathcal{O}(fg)$.
2) If c is a constant, then $c\mathcal{O}(g) = \mathcal{O}(g)$ and $\mathcal{O}(cg) = \mathcal{O}(g)$.
3) $\mathcal{O}(\mathcal{O}(g)) = \mathcal{O}(g)$.
4)

$$\sum_{i=k}^{m} \mathcal{O}(g^i) = \mathcal{O}(g^k), \qquad g \to 0, \qquad m \geq k$$

B. Asymptotic Behavior of the Second Moment

The following lemma uses the notion of functional order in asymptotic analysis to infer the asymptotic behavior of a contracting covariance matrix. The following definition will be useful.

Definition 3. The condition number of a matrix $A \in \mathbb{R}^{n \times m}$ is the ratio between its largest and smallest singular values: that is,

$$\kappa(A) \triangleq \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \tag{7}$$

Remark 1. The singular values of a symmetric positive definite matrix equal its eigenvalues.

Lemma 1. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix. The condition number of A satisfies

$$\kappa(A) \geq c(\operatorname{tr} A)^{-d} \quad \text{as } \operatorname{tr} A \to 0 \tag{8}$$

where $c > 0$ is a finite constant, if and only if there exists a quadratic form $\boldsymbol{g}^T A \boldsymbol{g} = \mathcal{O}((\operatorname{tr} A)^{d+1})$ as $\operatorname{tr} A \to 0$, for some unit vector $\boldsymbol{g}$ and a nonnegative scalar d.

Proof. (If.) Let G be an orthogonal matrix, and let $\boldsymbol{g}$ to be one of its columns. The quadratic form $\boldsymbol{g}^T A \boldsymbol{g}$ is a diagonal element of $G^T A G$. According to the Cauchy inclusion theorem, the minimal eigenvalue of $G^T A G$ is smaller than or equal to any eigenvalue of its submatrices, hence

$$\lambda_{\min}(G^T A G) \leq \boldsymbol{g}^T A \boldsymbol{g} = \mathcal{O}((\operatorname{tr} A)^{d+1}), \qquad \operatorname{tr} A \to 0 \tag{9}$$

The trace of a matrix equals the sum of its eigenvalues, hence

$$\lambda_{\max}(G^T A G) \geq \tfrac{1}{n} \operatorname{tr} A \tag{10}$$

Combining both Eqs. (9) and (10) yields

$$\kappa(G^T A G) \geq c(\operatorname{tr} A)^{-d} \quad \text{as } \operatorname{tr} A \to 0 \tag{11}$$

Notice that because G is orthogonal, both $G^T A G$ and A share the same eigenvalues (and hence the same condition number).

(Only if.) Let $\boldsymbol{g}_{\min}$ denote the unit eigenvector of A corresponding to $\lambda_{\min}(A)$. Then

$$\lambda_{\min}(A) = \boldsymbol{g}_{\min}^T A \boldsymbol{g}_{\min} \tag{12}$$

Because

$$\lambda_{\max}(A) \leq \operatorname{tr} A \tag{13}$$

it follows that

$$\lambda_{\max}(A) = \mathcal{O}(\operatorname{tr} A) \quad \text{as } \operatorname{tr} A \to 0 \tag{14}$$

It then follows from Eq. (8) that $\lambda_{\min}(A) = \mathcal{O}((\operatorname{tr} A)^{d+1})$ as $\operatorname{tr} A \to 0$. Thus, Eq. (12) yields

$$\boldsymbol{g}_{\min}^T A \boldsymbol{g}_{\min} = \mathcal{O}((\operatorname{tr} A)^{d+1}) \quad \text{as } \operatorname{tr} A \to 0 \tag{15}$$

thereby completing the proof.

Corollary 1. The smallest and largest eigenvalues of a matrix $A \in \mathbb{R}^{n \times n}$ satisfy

$$\lambda_{\min}(A) = \mathcal{O}(\lambda_{\max}(A)^{d+1}) \quad \text{as } \lambda_{\max}(A) \to 0 \tag{16}$$

if and only if A satisfies Lemma 1 for some nonnegative scalar d.

Proof. Because

$$\operatorname{tr} A \leq n\lambda_{\max}(A) \tag{17}$$

it follows that

$$\operatorname{tr} A = \mathcal{O}(\lambda_{\max}(A)) \quad \text{as } \lambda_{\max} \to 0 \tag{18}$$

Equations (14) and (18) render $\lambda_{\max}(A)$ and $\operatorname{tr} A$ interchangeable in terms of order symbols. Equations (16) and (8) thus imply each other straightforwardly. □

Consider now a constrained rv of the form

$$\boldsymbol{v} = [\boldsymbol{u}^T, u_n]^T, \qquad u_n = (\boldsymbol{u}^T \boldsymbol{u})^{(d+1)/2} \tag{19}$$

where d and $\boldsymbol{u}$ denote a nonnegative real scalar and a rv taking its values on $\mathbb{R}^{n-1}$, respectively. Let Σ be the second-moment matrix of $\boldsymbol{v}$. Using decomposition (19), the matrix Σ can be written as

$$\Sigma = E[\boldsymbol{v}\boldsymbol{v}^T] = \begin{bmatrix} E[\boldsymbol{u}\boldsymbol{u}^T] & E[(\boldsymbol{u}^T\boldsymbol{u})^{(d+1)/2}\boldsymbol{u}] \\ E[\boldsymbol{u}^T(\boldsymbol{u}^T\boldsymbol{u})^{(d+1)/2}] & E[(\boldsymbol{u}^T\boldsymbol{u})^{d+1}] \end{bmatrix} \tag{20}$$

Setting $\boldsymbol{g} = [\boldsymbol{0}_{1\times(n-1)}, 1]^T$, it follows that

$$\boldsymbol{g}^T \Sigma \boldsymbol{g} = E[(\boldsymbol{u}^T\boldsymbol{u})^{d+1}] \leq E[(\boldsymbol{u}^T\boldsymbol{u} + u_n^2)^{d+1}] = E[(\boldsymbol{v}^T\boldsymbol{v})^{d+1}] \tag{21}$$

Equation (21) implies that Σ satisfies Lemma 1 for the number $d \geq 0$ if

$$E[(\boldsymbol{v}^T\boldsymbol{v})^{d+1}] = \mathcal{O}((\operatorname{tr} \Sigma)^{d+1}), \qquad \operatorname{tr} \Sigma \to 0 \tag{22}$$

because then

$$\boldsymbol{g}^T \Sigma \boldsymbol{g} = \mathcal{O}((\operatorname{tr} \Sigma)^{d+1}), \qquad \operatorname{tr} \Sigma \to 0 \tag{23}$$

Equation (22) is a property of the distribution of the rv $\boldsymbol{v}$. The conditions for the existence of this property are detailed in the following proposition and in the ensuing discussion.

Proposition 1. Let $\boldsymbol{v}$ be a rv with a finite second-moment matrix Σ. Then Eq. (22) is satisfied for any $d \geq 0$ if

$$\sum_{i=1}^{\infty} i^{d+1} \operatorname{prob}((i-1)\operatorname{tr} \Sigma < \boldsymbol{v}^T\boldsymbol{v} \leq i \operatorname{tr} \Sigma) < \infty \tag{24}$$

where $\operatorname{prob}(C)$ denotes the probability of the event C.

The proof of Proposition 1 is deferred to Appendix A.

Corollary 2. Let $\boldsymbol{v}$ be a rv with finite second-moment matrix Σ, and let the pdf of $\boldsymbol{v}$ have a compact support.[‡] Then $\boldsymbol{v}$ satisfies Eq. (22).

Proof. From the finite cover property of compactness, it follows that there exists $N < \infty$ such that

[‡]A closed and bounded set in $\mathbb{R}^n$ is compact. A more general definition, which is more appealing in the context of Corollary 2, is the "covering" definition of compactness: A topological space is compact if each of its open covers has a finite subcover (a cover of a set X is a collection of sets such that X is a subset of the union of sets in the collection; a subcover of X is a subset of the cover of X that still covers X). The Heine–Borel theorem states that these two definitions of compactness are equivalent for subsets of Euclidean spaces. For further details, the reader is referred to [18].

$$\text{prob}\,((m-1)\text{tr}\,\Sigma < \boldsymbol{v}^T\boldsymbol{v} \le m\,\text{tr}\,\Sigma) = 0 \quad \forall\ m \ge N \tag{25}$$

which implies Eq. (24), thus yielding Eq. (22). □

III. Error Covariance of Quaternion Estimators

Forming the core of this work, this section applies the concepts developed in the previous section to the quaternion estimation error covariance matrix. The asymptotic behavior of this matrix is explored, and conclusions are drawn regarding the terms under which it can be adequately approximated.

A. Quaternion Estimators

The quaternion discrete-time process and observation equations can be described in a rather general form as

$$\mathbf{q}_k = \boldsymbol{\Phi}_k(\mathbf{q}_{k-1}, \boldsymbol{w}_k) \tag{26a}$$

$$\boldsymbol{y}_k = \boldsymbol{h}_k(\mathbf{q}_k, \boldsymbol{n}_k) \tag{26b}$$

where the quaternion rv $\mathbf{q}_k$ (respectively, its realization $\boldsymbol{q}_k$) takes values on the unit 3-sphere $\mathbb{S}^3$, and the observation rv $\boldsymbol{y}_k$ takes values in $\mathbb{R}^p$ (detailed expressions for the matrix functions $\boldsymbol{\Phi}_k$ and $\boldsymbol{h}_k$ can be found, for example, in [15]). The quaternion rv is written as

$$\mathbf{q} = \begin{bmatrix} \boldsymbol{\varrho} \\ q_4 \end{bmatrix} \tag{27}$$

where $\boldsymbol{\varrho}$ and q_4 denote its vector and scalar parts, respectively. In addition, $\{\boldsymbol{w}_k\}_{k=1}^{\infty}$ and $\{\boldsymbol{n}_k\}_{k=1}^{\infty}$ are the process and measurement white noises, respectively.

Let $\mathcal{Y}^k \triangleq \{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_k\}$ and $Y^k \triangleq \{\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_k\}$ be the measurement time history up to time k and its realization, respectively. In filtering algorithms, any statistical inference related to the unobserved signal (i.e., the state process) is based on the incoming measurements. Hence, the estimator of $\mathbf{q}_k$, denoted by $\hat{\mathbf{q}}_k$, is a random variable adapted to the filtration generated by the measurements up to time k (in other words, $\hat{\mathbf{q}}_k$ is $\mathcal{Y}^k$-measurable). In many cases, one is interested in obtaining the MMSE estimator of $\mathbf{q}_k$, which satisfies $\hat{\mathbf{q}}_k = E[\mathbf{q}_k|\mathcal{Y}^k]$. In general, though, $\hat{\mathbf{q}}_k$ may be any $\mathcal{Y}^k$-measurable function.

The ensuing development addresses sequential quaternion estimators. This class of estimators is characterized by the form

$$\hat{\mathbf{q}}_k = \mathcal{G}(\boldsymbol{y}_k, \hat{\mathbf{q}}_{k-1}) \tag{28}$$

where $\mathcal{G}(\cdot)$ is some measurable function.

B. Estimation Error Covariance

Let the additive estimation error belonging to the quaternion estimator $\hat{\mathbf{q}}_k$ be defined as

$$\tilde{\mathbf{q}}_k \triangleq \mathbf{q}_k - \hat{\mathbf{q}}_k \tag{29}$$

Notice that $\tilde{\mathbf{q}}_k$ is not necessarily a quaternion of rotation. Let P be the conditional covariance of the rotation quaternion estimation error: that is,

$$\begin{aligned} P &\triangleq \text{cov}(\tilde{\mathbf{q}}_k|\mathcal{Y}^k = Y^k) \\ &= E\{[\tilde{\mathbf{q}}_k - E(\tilde{\mathbf{q}}_k|\mathcal{Y}^k)][\tilde{\mathbf{q}}_k - E(\tilde{\mathbf{q}}_k|\mathcal{Y}^k)]^T|\mathcal{Y}^k = Y^k\} \end{aligned} \tag{30}$$

Using definition (29) in Eq. (30) yields

$$\begin{aligned} P = E\{&[\mathbf{q}_k - \hat{\mathbf{q}}_k - E(\mathbf{q}_k|\mathcal{Y}^k) + E(\hat{\mathbf{q}}_k|\mathcal{Y}^k)][\mathbf{q}_k - \hat{\mathbf{q}}_k - E(\mathbf{q}_k|\mathcal{Y}^k) \\ &+ E(\hat{\mathbf{q}}_k|\mathcal{Y}^k)]^T|\mathcal{Y}^k = Y^k\} \end{aligned} \tag{31}$$

Because $\hat{\mathbf{q}}_k$ is $\mathcal{Y}^k$-measurable (that is, $\hat{\mathbf{q}}_k$ is a function of the measurements $\mathcal{Y}^k$),

$$E(\hat{\mathbf{q}}_k|\mathcal{Y}^k) = \hat{\mathbf{q}}_k \tag{32}$$

Substituting Eq. (32) into Eq. (31) yields

$$P = E\{[\mathbf{q}_k - E(\mathbf{q}_k|\mathcal{Y}^k)][\mathbf{q}_k - E(\mathbf{q}_k|\mathcal{Y}^k)]^T|\mathcal{Y}^k = Y^k\} \tag{33}$$

which, by definition of the conditional covariance, is

$$P = \text{cov}(\mathbf{q}_k|\mathcal{Y}^k = Y^k) \tag{34}$$

so that the conditional covariances of both the quaternion and the quaternion estimation error are identical, independently of the estimator used.

Implying that the conditional estimation error covariances of all estimators are the same, the last observation might give rise to two questions:

1) How does this observation reconcile with the well-known fact that not all estimators are optimal for such a problem?

2) If (being equal for all estimators) the conditional covariance of the estimation error cannot be used to distinguish between different estimators, what is its role and why is it important in the estimation problem dealt with herein?

To answer these questions, notice that the MMSE estimator (which is the conditional expectation estimator) minimizes the mean square error (MSE) criterion:

$$\text{MSE} = \text{tr}\,P_m \tag{35}$$

where P_m is the estimation error second moment:

$$P_m \triangleq E\{(\mathbf{q}_k - \hat{\mathbf{q}}_k)(\mathbf{q}_k - \hat{\mathbf{q}}_k)^T\} \tag{36}$$

Furthermore, it can be shown that the conditional expectation estimator also minimizes P_m itself, in the sense that for any other $\mathcal{Y}^k$-measurable estimator $\hat{\mathbf{q}}_k$,[§]

$$E\{(\mathbf{q}_k - \hat{\mathbf{q}}_k)(\mathbf{q}_k - \hat{\mathbf{q}}_k)^T\} \succeq E\{(\mathbf{q}_k - E(\mathbf{q}_k|\mathcal{Y}^k))(\mathbf{q}_k - E(\mathbf{q}_k|\mathcal{Y}^k))^T\} \tag{37}$$

Using the smoothing property of the conditional expectation and explicitly writing the pdf with respect to which each expectation is carried out, Eq. (37) can be rewritten as

$$\begin{aligned} &E_{\mathcal{Y}}E_{\mathbf{q}_k|\mathcal{Y}}\{(\mathbf{q}_k - \hat{\mathbf{q}}_k)(\mathbf{q}_k - \hat{\mathbf{q}}_k)^T|\mathcal{Y}\} \\ &\quad\succeq E_{\mathcal{Y}}E_{\mathbf{q}_k|\mathcal{Y}}\{[\mathbf{q}_k - E(\mathbf{q}_k|\mathcal{Y}^k)][\mathbf{q}_k - E(\mathbf{q}_k|\mathcal{Y}^k)]^T|\mathcal{Y}\} \end{aligned} \tag{38}$$

which means that the MMSE estimator minimizes the conditional expectation of the second-moment matrix of the estimation error. Thus, the answer to the first question is that although the conditional covariance of the estimation error of all estimators is the same as that of the optimal MMSE estimator, only the MMSE estimator is optimal (in the sense of minimizing the MSE criterion), because optimality (in that sense) is determined by the conditional second-moment matrix of the estimation error, rather than by its conditional covariance.

Equation (38) also provides the answer to the second question, because it shows that the minimum (achieved by the MMSE estimator) of the conditional expectation of the estimation error second-moment matrix is the conditional covariance of $\mathbf{q}_k$: that is, P. Hence, the performance of any estimator designed to compute the MMSE estimate (or a practical approximation of it) can be assessed by the proximity of its computed conditional covariance to P.

The main result of this paper is presented in the following theorem.

Theorem 1. The following relations hold as $(\text{tr}\,P) \to 0$:

$$\kappa(P) \ge c_1(\text{tr}\,P)^{-1} \tag{39a}$$

$$\kappa(P) \ge c_2(\lambda_{\max}(P))^{-1} \tag{39b}$$

where $c_1, c_2 > 0$ are finite constants.

[§]Let A and B be two compatible matrices; then $A \succeq B$ if $A - B$ is positive semidefinite.

Proof. Let

$$\boldsymbol{\mu}_k \triangleq E[\mathbf{q}_k|\mathcal{Y}^k = Y^k] \tag{40}$$

and let $\boldsymbol{\gamma}_k \triangleq \boldsymbol{\mu}_k/\|\boldsymbol{\mu}_k\|$ be the normalized version of $\boldsymbol{\mu}_k$, then P can be rewritten as

$$P = E[(\mathbf{q}_k - \boldsymbol{\gamma}_k + \boldsymbol{\gamma}_k - \boldsymbol{\mu}_k)(\mathbf{q}_k - \boldsymbol{\gamma}_k + \boldsymbol{\gamma}_k - \boldsymbol{\mu}_k)^T|\mathcal{Y}^k = Y^k] \tag{41}$$

Expanding Eq. (41) yields

$$P = \tilde{P} - (1 - \|\boldsymbol{\mu}_k\|)^2\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T \tag{42}$$

where the matrix $\tilde{P}$ is defined as

$$\tilde{P} \triangleq E[(\mathbf{q}_k - \boldsymbol{\gamma}_k)(\mathbf{q}_k - \boldsymbol{\gamma}_k)^T|\mathcal{Y}^k = Y^k] \tag{43}$$

Let the matrix $\Psi \in \mathbb{R}^{4\times3}$ be chosen such that $\Gamma \triangleq [\Psi, \boldsymbol{\gamma}_k]$ is an orthogonal matrix, and let the rv $\boldsymbol{u}_k$ be defined as

$$\boldsymbol{u}_k \triangleq \Gamma^T\mathbf{q}_k - [\mathbf{0}_{1\times3}, 1]^T \tag{44}$$

Then

$$\begin{aligned}\Gamma^T P\Gamma &= \Gamma^T\tilde{P}\Gamma - \begin{bmatrix} 0_{3\times3} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{1\times3} & (1-\|\boldsymbol{\mu}_k\|)^2\end{bmatrix} \\ &= E[\boldsymbol{u}_k\boldsymbol{u}_k^T|\mathcal{Y}^k = Y^k] - \begin{bmatrix} 0_{3\times3} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{1\times3} & (1-\|\boldsymbol{\mu}_k\|)^2\end{bmatrix}\end{aligned} \tag{45}$$

Because Γ is an orthogonal matrix, $\Gamma^T\mathbf{q}_k$ is a rotation quaternion with a vector part $\boldsymbol{\varrho}_k = \Psi^T\mathbf{q}_k$. Using $\boldsymbol{\varrho}_k$, the vector $\boldsymbol{u}_k$ can be expressed as

$$\boldsymbol{u}_k = [\boldsymbol{\varrho}_k^T, (1 - \boldsymbol{\varrho}_k^T\boldsymbol{\varrho}_k)^{1/2} - 1]^T \tag{46}$$

Substituting Eq. (46) into Eq. (45) and extracting the fourth diagonal component of $\Gamma^T P\Gamma$ yields

$$\boldsymbol{\gamma}^T P\boldsymbol{\gamma} = E[((1 - \boldsymbol{\varrho}_k^T\boldsymbol{\varrho}_k)^{1/2} - 1)^2|\mathcal{Y}^k = Y^k] - (1 - \|\boldsymbol{\mu}_k\|)^2 \tag{47}$$

The fact that $\boldsymbol{\varrho}_k^T\boldsymbol{\varrho}_k \le 1$ facilitates the use of Lemma 2 (see Appendix C) in Eq. (47), thus

$$\begin{aligned}\boldsymbol{\gamma}^T P\boldsymbol{\gamma} &\le E[((1 - \boldsymbol{\varrho}_k^T\boldsymbol{\varrho}_k)^{1/2} - 1)^2|\mathcal{Y}^k = Y^k] \\ &\le E[(\boldsymbol{\varrho}_k^T\boldsymbol{\varrho}_k)^2|\mathcal{Y}^k = Y^k]\end{aligned} \tag{48}$$

The definition of Γ implies

$$\boldsymbol{\varrho}_k = \Psi^T(\mathbf{q}_k - \boldsymbol{\mu}_k) \tag{49}$$

Therefore, inequality (48) can be written as

$$\begin{aligned}\boldsymbol{\gamma}^T P\boldsymbol{\gamma} &\le E[((\mathbf{q}_k - \boldsymbol{\mu}_k)^T\Psi\Psi^T(\mathbf{q}_k - \boldsymbol{\mu}_k))^2|\mathcal{Y}^k = Y^k] \\ &= E[\|\Psi^T(\mathbf{q}_k - \boldsymbol{\mu}_k)\|^4|\mathcal{Y}^k = Y^k]\end{aligned} \tag{50}$$

Using Lemma 3 (see Appendix C), inequality (50) takes the form

$$\begin{aligned}\boldsymbol{\gamma}^T P\boldsymbol{\gamma} &\le E[((\mathbf{q}_k - \boldsymbol{\mu}_k)^T(\mathbf{q}_k - \boldsymbol{\mu}_k))^2|\mathcal{Y}^k = Y^k] \\ &= E[\|(\mathbf{q}_k - \boldsymbol{\mu}_k)\|^4|\mathcal{Y}^k = Y^k]\end{aligned} \tag{51}$$

Because the quaternion's distribution support is defined on the unit 3-sphere $\mathbb{S}^3$, it is compact, rendering the support of $(\mathbf{q}_k - \boldsymbol{\mu}_k)|\mathcal{Y}^k$ compact as well. Applying Corollary 2, Eq. (51) implies

$$\boldsymbol{\gamma}^T P\boldsymbol{\gamma} = \mathcal{O}((\mathrm{tr}\,P)^2) \quad \text{as tr}\,P \to 0 \tag{52}$$

According to Lemma 1, Eq. (52) implies Eq. (39a). Equation (39b) follows from Eq. (39a) upon observing that

$$\mathrm{tr}\ (P) \le 4\lambda_{\max}(P) \tag{53}$$

□

Theorem 1 immediately yields the following observation.

Corollary 3. The conditional covariance of the quaternion estimation error becomes asymptotically ill-conditioned as its trace tends to zero:

$$\kappa(P) \to \infty \quad \text{as tr}\,P \to 0 \tag{54}$$

C. Approximation of the Quaternion's Covariance

1. *Preliminaries: Statistical Taylor Expansions*

Let $\boldsymbol{v}$ and $\boldsymbol{V}$ be a rv taking values in $\mathbb{R}^n$ and its realization, respectively. Let also $f\colon \mathbb{R}^n \to \mathbb{R}$ be a C^∞ function. The multidimensional Taylor expansion of f about a nominal realization $\boldsymbol{V}^*$ is given by

$$f(\boldsymbol{v}) = \sum_{i=0}^{\infty}\frac{1}{i!}((\boldsymbol{v} - \boldsymbol{V}^*)^T\nabla_{\boldsymbol{u}})^i f(\boldsymbol{u})|_{\boldsymbol{u}=\boldsymbol{V}^*} \tag{55}$$

where the gradient operator is defined as

$$\nabla_{\boldsymbol{u}} \triangleq \left[\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_n}\right]^T \tag{56}$$

When using Eq. (55) to compute the statistics of a rv, the ith-order terms translate into moments of corresponding order. For instance, writing the expansion of $E[f(\boldsymbol{v})]$ yields

$$E[f(\boldsymbol{v})] = \sum_{i=0}^{\infty}\frac{1}{i!}E[((\boldsymbol{v} - \boldsymbol{V}^*)^T\nabla_{\boldsymbol{u}})^i f(\boldsymbol{u})|_{\boldsymbol{u}=\boldsymbol{V}^*}] \tag{57}$$

which is a sum of moments of various orders of elements of $\boldsymbol{v}$.

Proposition 2. Let $\boldsymbol{v}$ and $\boldsymbol{V}^*$ be a rv taking values in $\mathbb{R}^n$ and its nominal realization, respectively. Suppose that $\boldsymbol{v} - \boldsymbol{V}^*$ satisfies Proposition 1 for any number $d \ge 0$. Then, for any sufficiently differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$, the mth-order Taylor expansion of $\varphi(\boldsymbol{v}) \triangleq E[f(\boldsymbol{v})]$ about $\boldsymbol{V}^*$ satisfies

$$\hat{\varphi}_{[m]} = \sum_{i=0}^{m}\mathcal{O}((\mathrm{tr}\,D)^{i/2}) \quad \text{as tr}\,D \to 0 \tag{58}$$

where $D \triangleq E[(\boldsymbol{v} - \boldsymbol{V}^*)(\boldsymbol{v} - \boldsymbol{V}^*)^T]$.

The proof of Proposition 2 is deferred to Appendix B.

2. *Approximate Covariance Matrix*

Conveyed by Theorem 1, the asymptotic properties of the quaternion's estimation error covariance matrix, P, do not carry over to low-order Taylor expansions of P. This claim is made precise and established in the ensuing.

Definition 4. The mth-order Taylor expansion of the quaternion estimation error covariance P about some nominal quaternion realization $\boldsymbol{q}_k^*$ is the 4×4 matrix $\hat{P}_{[m]}$ for which the elements are the corresponding mth-order Taylor expansions about $\boldsymbol{q}_k^*$ of the corresponding elements of P.

Consider the quadratic form

$$\varphi(\mathbf{q}_k) \triangleq \boldsymbol{g}^T P\boldsymbol{g} \tag{59}$$

where P is defined in Eq. (30) and $\boldsymbol{g} \in \mathbb{R}^4$ is a unit vector. Recognizing that

$$\varphi(\mathbf{q}_k) = E[(\boldsymbol{g}^T(\mathbf{q}_k - E[\mathbf{q}_k|\mathcal{Y}^k = Y^k]))^2|\mathcal{Y}^k = Y^k] \tag{60}$$

the corresponding mth-order Taylor expansion of $\varphi(\mathbf{q}_k)$ about some nominal quaternion $\mathbf{q}_k^*$ is written as

$$\hat{\varphi}_{[m]} = \sum_{i=0}^{m}\frac{1}{i!}E[((\mathbf{q}_k - \boldsymbol{q}_k^*)^T\nabla_{\mathbf{q}_k})^i f(\mathbf{q}_k, Y^k)|_{\mathbf{q}_k=\boldsymbol{q}_k^*}|\mathcal{Y}^k = Y^k] \tag{61}$$

where the sufficiently differentiable function $f(\mathbf{q}_k, Y^k)$ is defined as

$$f(\mathbf{q}_k, Y^k) \triangleq (\boldsymbol{g}^T(\mathbf{q}_k - E[\mathbf{q}_k|\mathcal{Y}^k = Y^k]))^2 \tag{62}$$

Because $\boldsymbol{g}$ is a constant vector, it can be deduced that

$$\boldsymbol{g}^T\hat{P}_{[m]}\boldsymbol{g} = \hat{\varphi}_{[m]} \tag{63}$$

where $\hat{P}_{[m]}$ denotes the mth-order Taylor expansion of P.

Proposition 3. Let $\hat{P}_{[m]}$ be the mth-order Taylor expansion of the quaternion estimation error covariance P about some nominal quaternion realization $\boldsymbol{q}_k^*$. Assume that $\hat{P}_{[m]}$ consists of second- and higher-order terms only: that is, for any unit vector $\boldsymbol{g} \in \mathbb{R}^4$ it satisfies

$$\boldsymbol{g}^T\hat{P}_{[m]}\boldsymbol{g} = \sum_{i=2}^{m}\frac{1}{i!}E[((\mathbf{q}_k - \boldsymbol{q}_k^*)^T\nabla_{\mathbf{q}_k})^i f(\mathbf{q}_k, Y^k)|_{\mathbf{q}_k=\boldsymbol{q}_k^*}|\mathcal{Y}^k = Y^k] \tag{64}$$

where the sufficiently differentiable function $f(\mathbf{q}_k, Y^k)$ is defined in Eq. (62). Then Theorem 1 does not apply to the second- and third-order Taylor approximations of P, $\hat{P}_{[2]}$, and $\hat{P}_{[3]}$, respectively.

Proof. Because the quaternion rv has a compact support, it satisfies Proposition 1 for any number $d \geq 0$. Thus, using Proposition 2 yields

$$\boldsymbol{g}^T\hat{P}_{[m]}\boldsymbol{g} = \mathcal{O}((\operatorname{tr} D)^{M/2}) \quad \text{as } \operatorname{tr} D \to 0 \tag{65}$$

where $D \triangleq E[(\mathbf{q}_k - \boldsymbol{q}_k^*)(\mathbf{q}_k - \boldsymbol{q}_k^*)^T|\mathcal{Y}^k = Y^k]$, and the integer M, satisfying $2 \leq M \leq m$, depends on the selection of $\boldsymbol{g}$. Hence, using the least possible order of tr$D$ (i.e., $M = 2$) gives

$$\operatorname{tr} \hat{P}_{[m]} = \sum_{j=1}^{4}\boldsymbol{e}_j^T\hat{P}_{[m]}\boldsymbol{e}_j = \mathcal{O}(\operatorname{tr} D) \quad \text{as } \operatorname{tr} D \to 0 \tag{66}$$

where $\boldsymbol{e}_j$ is a unit vector having 1 as its jth element.

Considering Theorem 1, for $\hat{P}_{[m]}$ to satisfy

$$\kappa(\hat{P}_{[m]}) \geq c(\operatorname{tr}\hat{P}_{[m]})^{-1} \quad \text{as } \operatorname{tr}\hat{P}_{[m]} \to 0 \tag{67}$$

Lemma 1 states that the following condition must hold,

$$\boldsymbol{g}^T\hat{P}_{[m]}\boldsymbol{g} = \mathcal{O}((\operatorname{tr}\hat{P}_{[m]})^2) \quad \text{as } \operatorname{tr}\hat{P}_{[m]} \to 0 \tag{68}$$

for some unit vector $\boldsymbol{g}$. Using both Eqs. (65) and (66), Eq. (68) translates into

$$\mathcal{O}((\operatorname{tr} D)^{M/2}) = \mathcal{O}((\operatorname{tr} D)^2) \quad \text{as } \operatorname{tr} D \to 0, \qquad 2 \leq M \leq m \tag{69}$$

Equation (69) can be satisfied for $M \geq 4$ only, implying that low-order approximations of P, corresponding to $m < 4$, do not comply with Theorem 1. □

Corollary 4. Approximations of P of orders lower than 4 do not share their asymptotic properties with P.

Proof. The proof follows straightforwardly from Theorem 1 and Proposition 3.

D. Illustrative 2-D Example

Although Corollary 4 addresses the quaternion covariance matrix, it is conceivable that this claim applies to other matrices (of any order) possessing similar properties as well. Hence, it is interesting to examine the asymptotic behavior of a corresponding low-order Taylor expansion of the covariance matrix of the rv of Eq. (1).

Because $C \triangleq \operatorname{cov}(\boldsymbol{x})$ is a diagonal matrix, its second-order Taylor expansion about some $a^* \neq 0$ is given by $\hat{C}_{[2]} = \operatorname{diag}\{\hat{\lambda}_1, \hat{\lambda}_2\}$, where

$$\hat{\lambda}_i = \lambda_i|_{a=a^*} + \frac{\partial\lambda_i}{\partial a}|_{a=a^*}(a - a^*) + \frac{1}{2}\frac{\partial^2\lambda_i}{\partial a^2}|_{a=a^*}(a - a^*)^2, \qquad i = 1, 2 \tag{70}$$

and

$$\frac{\partial\lambda_1}{\partial a} = \frac{1}{2}\frac{\cos a}{a} - \frac{1}{2}\frac{\sin a}{a^2} - 2\frac{\sin a}{a^2} + 8\frac{\sin^2(a/2)}{a^3} \tag{71a}$$

$$\frac{\partial\lambda_2}{\partial a} = -\frac{1}{2}\frac{\cos a}{a} + \frac{1}{2}\frac{\sin a}{a^2} \tag{71b}$$

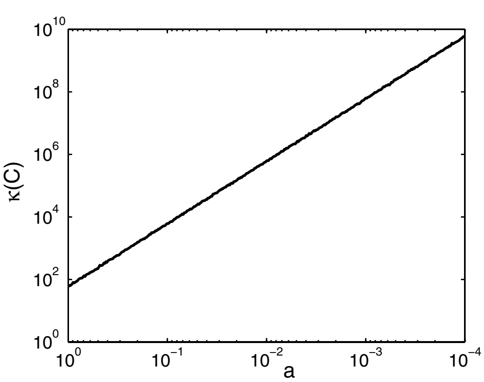
$$\frac{\partial^2\lambda_1}{\partial a^2} = -\frac{1}{2}\frac{\sin a}{a} - \frac{\cos a}{a^2} + \frac{\sin a}{a^3} - 2\frac{\cos^2(a/2)}{a^2} + 8\frac{\sin a}{a^3} + 2\frac{\sin^2(a/2)}{a^2} - 24\frac{\sin^2(a/2)}{a^4} \tag{71c}$$

$$\frac{\partial^2\lambda_2}{\partial a^2} = \frac{1}{2}\frac{\sin a}{a} + \frac{\cos a}{a^2} - \frac{\sin a}{a^3} \tag{71d}$$

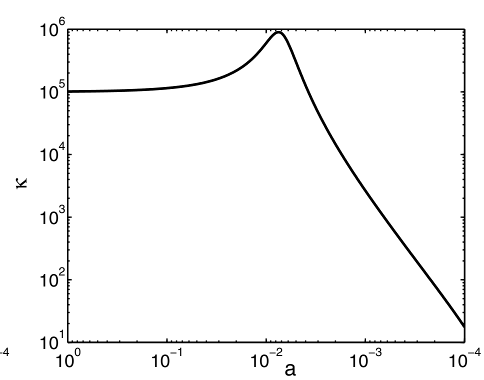
The condition numbers of both the true and the (second-order) approximate covariance matrices were numerically computed using Eqs. (2), (70), and (71). The behavior of both $\kappa(C)$ and $\kappa(\hat{C}_{[2]})$ as $a \to 0$ is depicted in Fig. 1. Unsurprisingly, as predicted by Corollary 4, the condition numbers of both matrices exhibit entirely different asymptotic behavior.

E. Discussion

The preceding analysis shows that the quaternion estimation error covariance matrix does possess the unique property of becoming ill-conditioned as its trace tends to zero (see Corollary 3). However, this property is not inherited by its corresponding second- and third-order Taylor approximations (see Corollary 4). Recalling that the EKF-computed covariance matrix consists of second-order terms only (see Appendix D for an overview of the EKF mechanization), the analysis presented herein explains why the AEKF has never been shown, in practice, to compute an ill-conditioned covariance matrix.



a) $\kappa(C)$



b) $\kappa(\hat{C}_{[2]})$

Fig. 1 Asymptotic condition number of the true (left panel) and (second-order) approximate (right panel) covariance of $\boldsymbol{x}$.

Recently, several conditions were given in [14,19] for the EKF-computed covariance matrix to become ill-conditioned or even singular. Both of these works suggest that the distinctive feature that enables covariance ill-conditioning is the absence of process noise. This claim is investigated using a numerical example in Sec. V.

IV. Geometrical Interpretation

Invoking the concept of smooth manifolds from differential geometry reveals some interesting geometrical meanings of the previously obtained results. Manifolds are topological spaces in which the neighborhood of each point resembles an open subset of $\mathbb{R}^n$. The dimension of the Euclidean space in the vicinity of a point on the manifold indicates the manifold's dimension. The 3-dimensional sphere, for instance, is a 2-dimensional smooth manifold (denoted as $\mathbb{S}^2$ or 2-sphere), implying that to within a small region on its surface, every point lies on a 2-D plane. The latter insight suggests that the third dimension of a contracting domain on the 2-sphere vanishes faster than the other two.

A. Contraction on a Smooth Manifold

Let M be an r-dimensional smooth manifold in $\mathbb{R}^n$ $(n > r)$, and let $\boldsymbol{\phi}\colon M \to \mathbb{R}^r$ denote some coordinate chart. Let also $\tilde{P}$ be the second-moment matrix of the difference rv:

$$\mathrm{d}\boldsymbol{v} \triangleq \boldsymbol{v} - \boldsymbol{V}_0 \tag{72}$$

where $\boldsymbol{v}$ denotes a rv taking its values on M, and $\boldsymbol{V}_0 \in M$ is some deterministic point. In the case under consideration, the asymptotic characteristics of $\tilde{P}$ refer to the behavior of its eigenvalues as $\operatorname{tr}\tilde{P} \to 0$. Bearing in mind that the support of the probability distribution of $\boldsymbol{v}$ contracts to a single point on M, the notion of manifold tangent space becomes useful for examining the eigenvalues of $\tilde{P}$.

Let $\boldsymbol{V}$ be a realization of the rv $\boldsymbol{v}$. Let also $\boldsymbol{x} \triangleq [x_1, \ldots, x_r]^T$ and $\boldsymbol{X} \triangleq [X_1, \ldots, X_r]^T$ denote a rv taking values on $\mathbb{R}^r$ and its realization, respectively. In the vicinity of $\boldsymbol{V}_0$ the difference in Eq. (72) is just an infinitesimal variation of $\boldsymbol{v}$. Thus, using the chain rule of differentials and the coordinate chart $\boldsymbol{\phi}(\cdot)$, it follows that

$$\mathrm{d}\boldsymbol{V} = J(\boldsymbol{X})\,\mathrm{d}\boldsymbol{X}, \qquad \operatorname{tr}\tilde{P} \to 0 \tag{73}$$

where

$$J(\boldsymbol{X}) \triangleq \frac{\partial \boldsymbol{\phi}^{-1}(\boldsymbol{X})}{\partial \boldsymbol{X}} \in \mathbb{R}^{n\times r} \tag{74}$$

Now let $\boldsymbol{c} \triangleq \mathrm{d}\boldsymbol{X}/\mathrm{d}X$, where $\mathrm{d}X \triangleq \|\mathrm{d}\boldsymbol{X}\|$. Using this definition, Eq. (73) is rewritten as

$$\mathrm{d}\boldsymbol{V} = J(\boldsymbol{X})\boldsymbol{c}\,\mathrm{d}X, \qquad \operatorname{tr}\tilde{P} \to 0 \tag{75}$$

which expresses the fact that $\mathrm{d}\boldsymbol{V}$ can be written as a linear combination of vectors [the columns of $J(\boldsymbol{X})$] in the tangent space $T(M)$.

Letting $\boldsymbol{g} \in \mathbb{R}^n$ be some unit-norm vector, Eq. (75) yields

$$\boldsymbol{g}^T\tilde{P}\boldsymbol{g} = E[(\boldsymbol{g}^T J(\boldsymbol{x}) G J(\boldsymbol{x})^T \boldsymbol{g})(\mathrm{d}x)^2], \qquad \operatorname{tr}\tilde{P} \to 0 \tag{76}$$

where $G \triangleq \boldsymbol{c}\boldsymbol{c}^T$. Equation (76) implies

$$\min_{\boldsymbol{\phi}^{-1}(X)\in M} (\boldsymbol{g}^T J(\boldsymbol{X}) G J(\boldsymbol{X})^T \boldsymbol{g}) \le \frac{\boldsymbol{g}^T\tilde{P}\boldsymbol{g}}{E[(\mathrm{d}x)^2]} \le \max_{\boldsymbol{\phi}^{-1}(X)\in M} (\boldsymbol{g}^T J(\boldsymbol{X}) G J(\boldsymbol{X})^T \boldsymbol{g}) \tag{77}$$

as $\operatorname{tr}\tilde{P} \to 0$. Now the condition number of $\tilde{P}$ satisfies

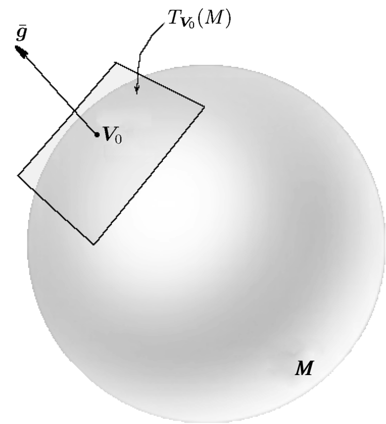


Fig. 2 **$T_{V_0}(M)$ is the tangent space of the manifold M at V_0. The vector $\bar{g}$ is in the cotangent space $T_{V_0}^{\perp}(M)$. If, for example, M is the 2-sphere, then the tangent space is a 2-dimensional plane.**

$$\kappa(\tilde{P}) = \frac{\max_{g} \boldsymbol{g}^T\tilde{P}\boldsymbol{g}}{\min_{g} \boldsymbol{g}^T\tilde{P}\boldsymbol{g}} \ge \frac{\max_{g} \min_{\boldsymbol{\phi}^{-1}(X)\in M} (\boldsymbol{g}^T J(\boldsymbol{X}) G J(\boldsymbol{X})^T \boldsymbol{g})}{\min_{g} \max_{\boldsymbol{\phi}^{-1}(X)\in M} (\boldsymbol{g}^T J(\boldsymbol{X}) G J(\boldsymbol{X})^T \boldsymbol{g})}, \qquad \operatorname{tr}\tilde{P} \to 0 \tag{78}$$

Taking the limit in Eq. (78), the support of the distribution of $\boldsymbol{v}$ contracts to a single point having the coordinate $\boldsymbol{X}_0 = \boldsymbol{\phi}(\boldsymbol{V}_0)$, thereby yielding

$$\lim_{\operatorname{tr}\tilde{P}\to 0} \kappa(\tilde{P}) \ge \frac{\max_{g}(\boldsymbol{g}^T J(\boldsymbol{X}_0) G J(\boldsymbol{X}_0)^T \boldsymbol{g})}{\min_{g}(\boldsymbol{g}^T J(\boldsymbol{X}_0) G J(\boldsymbol{X}_0)^T \boldsymbol{g})} \tag{79}$$

Now the dimension theorem yields

$$\dim\ker J(\boldsymbol{X}_0)^T \ge n - r \tag{80}$$

implying that there exists a unit-norm vector $\bar{\boldsymbol{g}} \in \mathbb{R}^n$ for which $J(\boldsymbol{X}_0)^T\bar{\boldsymbol{g}} = \boldsymbol{0}_{r\times 1}$. In turn, this implies that the denominator in Eq. (79) equals zero, yielding

$$\kappa(\tilde{P}) \to \infty \qquad \text{as } \operatorname{tr}\tilde{P} \to 0 \tag{81}$$

irrespective of $\boldsymbol{V}_0$.

The preceding argumentation renders $\bar{\boldsymbol{g}}$ the direction associated with the maximal contraction as $\operatorname{tr}\tilde{P} \to 0$. Let $\boldsymbol{u}_i(\boldsymbol{X}_0) \in \mathbb{R}^n$ $(i = 1, \ldots, r)$ be the columns of $J(\boldsymbol{X}_0)$, and let $T_{V_0}(M)$ be the tangent space of M at $\boldsymbol{V}_0$. Then

$$T_{V_0}(M) = \operatorname{span}(\{\boldsymbol{u}_i(\boldsymbol{X}_0)\}_{i=1}^r) \tag{82}$$

Because $\boldsymbol{u}_i(\boldsymbol{X}_0)^T\bar{\boldsymbol{g}} = 0$ $(i = 1, \ldots, r)$, it follows that

$$\bar{\boldsymbol{g}} \in T_{V_0}^{\perp}(M), \qquad \operatorname{tr}\tilde{P} \to 0 \tag{83}$$

where $T_{V_0}^{\perp}(M)$ denotes the cotangent space of M at $\boldsymbol{V}_0$. In other words, the maximal contraction occurs in a direction perpendicular to the tangent space at $\boldsymbol{V}_0$ (see Fig. 2).

B. Orthogonality Principle on $\mathbb{S}^3$

Because the rotation quaternion is defined on the unit 3-sphere $\mathbb{S}^3$, the preceding insights apply to the quaternion covariance matrix P as well.[¶] Thus, the condition number of P grows without bound as its trace tends to zero. Furthermore, Eq. (52) suggests that in this case, the maximal contraction occurs in the direction of the quaternion's MMSE estimate. Hence, according to Eq. (83), the MMSE estimate

[¶]The minor difference between the covariance matrix P and the second-moment matrix $\tilde{P}$ is pointed out by Eq. (42) in Sec. III. It is further shown in the same section that both matrices share the same asymptotic behavior.

becomes perpendicular to the tangent space of $\mathbb{S}^3$ at the contraction point.

Further assuming that $\hat{\boldsymbol{q}}$ is the normalized MMSE estimator of $\boldsymbol{q}$ renders $\tilde{P}$ the quaternion estimation error second-moment matrix [see Eq. (43)]. Both Eqs. (72) and (75) imply that the asymptotic quaternion estimation error is a linear combination of vectors in the tangent space at a specific point on $\mathbb{S}^3$, which consequently means that the estimation error becomes asymptotically perpendicular to the MMSE quaternion estimate. This asymptotic property concurs with the well-known orthogonality principle associated with Euclidean MMSE estimators (see [20], p. 177).

V. Simulation Study

A. Quaternion Particle Filter

A simulation study has been carried out to demonstrate the conclusions of previous sections. The asymptotic behavior of the quaternion's estimation error covariance matrix, stated in Theorem 1, was examined using the quaternion particle filter (QPF) of [15]. The QPF is a quaternion-based particle filter that approximates the posterior filtering density by means of samples (particles). This type of estimator is capable of capturing the statistics up to any order, at the expense of computational efficiency. The QPF was applied to a conventional attitude estimation problem, assuming the following measurement model:

$$\boldsymbol{b}_k = A_k \boldsymbol{r}_k + \boldsymbol{n}_k, \qquad \boldsymbol{n}_k \sim \mathcal{N}(\mathbf{0}_{3\times 1}, R) \tag{84}$$

where $\boldsymbol{b}_k$ and $\boldsymbol{r}_k$ denote body-fixed and reference-frame vectors, respectively. The attitude matrix at time k, denoted by A_k, is related to the quaternion via the expression

$$A_k \triangleq A(\mathbf{q}_k) = [(\mathrm{q}_{4_k})^2 - \boldsymbol{\varrho}_k^T \boldsymbol{\varrho}_k] I_{3\times 3} + 2\boldsymbol{\varrho}_k \boldsymbol{\varrho}_k^T - 2\mathrm{q}_{4_k}[\boldsymbol{\varrho}_k \times] \tag{85}$$

where $[\boldsymbol{a}\times]$ denotes the cross-product matrix associated with $\boldsymbol{a}$: that is,

$$[\boldsymbol{a}\times] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \tag{86}$$

The process noise is a zero-mean white noise with intensity 0.94 $(\mu\text{rad})^2/\text{s}$, which is injected through the quaternion transition matrix.

The results of the simulation runs corroborate the predictions of Theorem 1 regarding the behavior of the estimation error covariance matrix of the QPF. Figure 3 shows the eigenvalues and the condition number of the covariance matrix as computed by the QPF algorithm using 5000 particles in a typical run. The diagonal straight line in Fig. 3b corresponds to $1/\lambda_{\max}$. From this figure it is evident that the condition number of the covariance matrix is of the order of $1/\lambda_{\max}$, as $\lambda_{\max} \to 0$. In passing, it is noted that similar results have been obtained using only 150 particles (which demonstrates the efficiency and accuracy of the QPF algorithm).

B. Additive EKF

To demonstrate Corollary 4, the covariance matrix of the AEKF algorithm of [11] was computed. As in the QPF case, the AEKF was applied to a conventional attitude estimation problem, in which the same vector measurement model was used. The AEKF initial conditions and the noise statistical properties were set in accordance with the conclusions in [14,19]. The authors of [14,19] conclude that the AEKF covariance matrix becomes ill-conditioned, and even singular, in one of the following situations: 1) in the absence of process noise, 2) when the attitude estimation error covariance becomes very small, and 3) when the initial covariance matrix reflects extremely high certainty in the fourth quaternion estimation error component. In other words, when P_0 is ill-conditioned.

The AEKF was extensively tested under the aforementioned conditions, using 5000 Monte Carlo runs. In each run, the AEKF was implemented with the exclusion of process noise. The measurement noise covariance was set to $R = 10^{-11} I_{3\times 3}$, allowing the attitude covariance to reach extremely small values within a short time. The initial covariance condition number was set within the range of 10^6 to 10^8, where the first 3 diagonal elements of P_0 were uniformly sampled over the interval $[10^{-7}, 10^{-5}]$. In each run, the true initial attitude quaternion $\boldsymbol{q}_0$ was uniformly sampled on the unit 3-sphere, and the initial estimate was set as

$$\hat{\boldsymbol{q}}_0 = \boldsymbol{X}_0 / \|\boldsymbol{X}_0\|, \qquad \boldsymbol{X}_0 \sim \mathcal{N}(\boldsymbol{q}_0, P_0) \tag{87}$$

Figure 4 shows the distributions (via their percentile lines) of the maximal eigenvalue and the condition number of the AEKF-computed covariance matrix. This figure clearly validates Corollary 4. The AEKF-computed covariance does not exhibit the asymptotic properties conveyed by Theorem 1 and, except for a brief transient at the beginning of the runs, its condition number is "well-behaved" and improves with time.

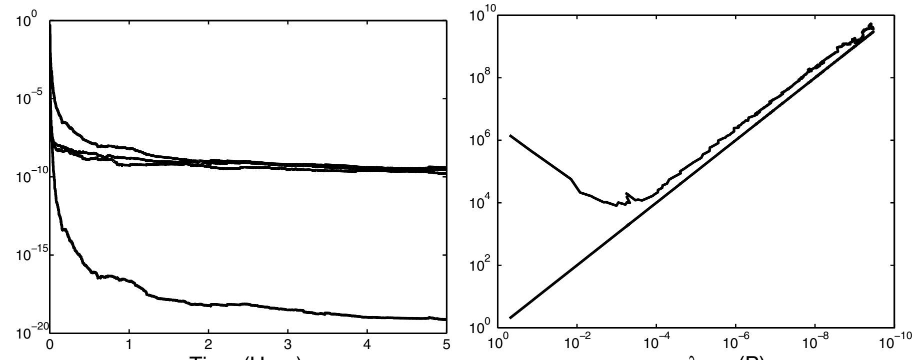
C. AEKF in a Static Case

Consider the problem of estimating the attitude of a nonrotating body (i.e., having zero dynamics and no process noise) using vector observations. In this case, the AEKF covariance update equation can be expressed in information form as

$$P_{k+1/k+1}^{-1} = P_{k/k}^{-1} + H_k^T R^{-1} H_k \tag{88}$$

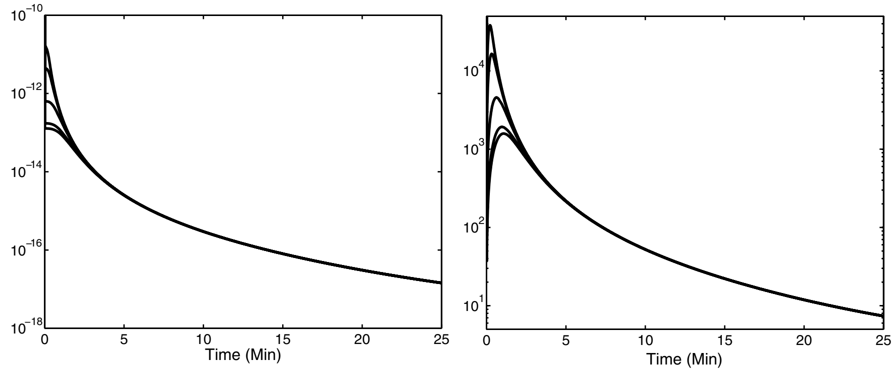
where the 3×4 measurement sensitivity matrix H_k depends on both the quaternion estimate $\hat{\mathbf{q}}_k$ and a reference vector measurement $\boldsymbol{r}_k$: that is [11],

$$H_k \triangleq H(\hat{\mathbf{q}}_k, \boldsymbol{r}_k) = \left[\left. \frac{\partial [A(\boldsymbol{q})\boldsymbol{r}_k]^T}{\partial \boldsymbol{q}} \right|_{\boldsymbol{q}=\hat{\mathbf{q}}_k} \right]^T \tag{89}$$



a) $\lambda_i(P)$, $i = 1, 2, 3, 4$ b) $\kappa(P)$

Fig. 3 Eigenvalues (left panel) and condition number (right panel) of the quaternion estimation error covariance matrix as computed by the QPF algorithm using 5000 particles. The diagonal straight line in the right panel corresponds to $1/\lambda_{\max}$.



a) $\lambda_{\max}(\hat{\boldsymbol{P}})$ **b)** $\kappa(\hat{\boldsymbol{P}})$

Fig. 4 Distribution of maximal eigenvalue (left panel) and condition number (right panel) of AEKF-computed covariance matrix in 5000 Monte Carlo runs. Lines, top to bottom, correspond to 95, 85, 50, 15, and 5 percentiles.

Suppose that the attitude estimation errors are small, such that

$$\tilde{\mathbf{q}}_k = \hat{\mathbf{q}}_k - \boldsymbol{q} = [\mathcal{O}(\epsilon), \mathcal{O}(\epsilon), \mathcal{O}(\epsilon), \mathcal{O}(\epsilon)]^T \quad \text{with probability 1}, \quad \epsilon \to 0 \tag{90}$$

where $\boldsymbol{q} \triangleq [q_1,\ q_2,\ q_3,\ q_4]^T$ denotes the true attitude quaternion. Using Eq. (90) in Eq. (89) while retaining the least-order term of the corresponding Taylor expansion of H_k about $\boldsymbol{q}$ yields

$$H_k = H(\boldsymbol{q}, \boldsymbol{r}_k) + \Delta H(\boldsymbol{q}, \boldsymbol{r}_k) \tag{91}$$

where the entries of the 3×4 random matrix $\Delta H(\boldsymbol{q}, \boldsymbol{r}_k)$ satisfy

$$\Delta H_{ij}(\boldsymbol{q}, \boldsymbol{r}_k) = (\nabla_u H_{ij}(\boldsymbol{u}, \boldsymbol{r}_k)|_{\boldsymbol{u}=\boldsymbol{q}})^T \tilde{\mathbf{q}}_k = \mathcal{O}(\epsilon) \quad \text{with probability 1}, \quad \epsilon \to 0 \tag{92}$$

Define

$$C_k \triangleq \frac{1}{k} P_{k|k}^{-1} \tag{93}$$

Using Eq. (88), C_k can be expressed as

$$\begin{aligned} C_k &= \frac{1}{k}\left[P_{0|0}^{-1} + \sum_{i=1}^{k} H_i^T R^{-1} H_i\right] \\ &= \frac{1}{k} P_{0|0}^{-1} + \frac{1}{k}\sum_{i=1}^{k} H(\boldsymbol{q}, \boldsymbol{r}_i)^T R^{-1} H(\boldsymbol{q}, \boldsymbol{r}_i) \\ &+ \frac{1}{k}\sum_{i=1}^{k} H(\boldsymbol{q}, \boldsymbol{r}_i)^T R^{-1} \Delta H(\boldsymbol{q}, \boldsymbol{r}_i) \\ &+ \frac{1}{k}\sum_{i=1}^{k} \Delta H(\boldsymbol{q}, \boldsymbol{r}_i)^T R^{-1} H(\boldsymbol{q}, \boldsymbol{r}_i) \\ &+ \frac{1}{k}\sum_{i=1}^{k} \Delta H(\boldsymbol{q}, \boldsymbol{r}_i)^T R^{-1} \Delta H(\boldsymbol{q}, \boldsymbol{r}_i) \end{aligned} \tag{94}$$

Now assume that $\{\boldsymbol{r}_i\}_{i=1}^k$ is an independent and identically distributed sequence of sample vectors drawn from the probability distribution of the rv $\boldsymbol{r}$. Then, from the strong law of large numbers, it follows that

$$\lim_{k\to\infty} C_k = S + S_1 + S_2 \quad \text{with probability 1} \tag{95}$$

with S, S_1, and S_2 given by

$$S = E[H(\boldsymbol{q}, \boldsymbol{r})^T R^{-1} H(\boldsymbol{q}, \boldsymbol{r})] \tag{96a}$$

$$S_1 = E[H(\boldsymbol{q}, \boldsymbol{r})^T R^{-1} \Delta H(\boldsymbol{q}, \boldsymbol{r})] + E[\Delta H(\boldsymbol{q}, \boldsymbol{r})^T R^{-1} H(\boldsymbol{q}, \boldsymbol{r})] \tag{96b}$$

$$S_2 = E[\Delta H(\boldsymbol{q}, \boldsymbol{r})^T R^{-1} \Delta H(\boldsymbol{q}, \boldsymbol{r})] \tag{96c}$$

where all expectations are performed with respect to the random vector $\boldsymbol{r}$. Using Lemma 4 (see Appendix C) in Eqs. (96b) and (96c) yields

$$(S_1)_{ij} = \mathcal{O}(\epsilon), \qquad (S_2)_{ij} = \mathcal{O}(\epsilon^2), \qquad \epsilon \to 0 \tag{97}$$

Equations (93), (95), and (97) yield

$$\lim_{k\to\infty} C_k = \lim_{k\to\infty} (k P_{k|k})^{-1} = \bar{S} \quad \text{with probability 1} \tag{98}$$

where $\bar{S}_{ij} \triangleq S_{ij} + \mathcal{O}(\epsilon)$ as $\epsilon \to 0$. Perhaps counterintuitively, the rank deficiency of the matrix $H(\boldsymbol{q}, \boldsymbol{r})^T R^{-1} H(\boldsymbol{q}, \boldsymbol{r})$ does not necessarily imply singularity of S in Eq. (96a). In fact, considering the random nature of the vector observation $\boldsymbol{r}$, this possibility can be outright rejected, as established in Lemma 5 (see Appendix C).

Now, because the eigenvalues of a matrix are continuous functions of its entries, it follows from Eq. (98) that

$$\lambda_i(\bar{S}) = \lim_{k\to\infty} \frac{1}{k} \lambda_i(P_{k|k}^{-1}) \quad \text{with probability 1}, \quad i = 1, \ldots, 4 \tag{99}$$

Also, $\lambda_{\min}(\bar{S}) \approx \lambda_{\min}(S)$, and Lemma 5 states that $\lambda_{\min}(S) > 0$, hence Eq. (99) yields

$$\begin{aligned} \kappa(\bar{S}) &= \frac{\lim_{k\to\infty} \frac{1}{k} \lambda_{\max}(P_{k|k}^{-1})}{\lim_{k\to\infty} \frac{1}{k} \lambda_{\min}(P_{k|k}^{-1})} = \lim_{k\to\infty} \kappa(P_{k|k}^{-1}) \\ &= \lim_{k\to\infty} \kappa(P_{k|k}) \quad \text{with probability 1} \end{aligned} \tag{100}$$

Thus, under the assumptions of the specific example presented here, the asymptotic condition number of the AEKF-computed covariance should equal the condition number of $\bar{S} \approx S$ almost surely. The matrix S can be easily approximated via the Monte Carlo sampling method. Using 5×10^4 vector samples corresponding to the uniform distribution

$$\boldsymbol{r} \sim [U(-30, 30), U(-30, 30), U(-30, 30)]^T \tag{101}$$

and setting $R = 10^{-11} I_{3\times3}$, the condition number of the computed S matrix was found to be $\kappa(S) = 1.5028$. On the other hand, the limit condition number $\lim_{k\to\infty} \kappa(P_{k|k})$ was evaluated using a Monte Carlo simulation consisting of 1000 AEKF runs. In all runs, the initial attitude estimation error was set as described in Sec. V.B. Figure 5 shows the distribution of $\kappa(P_{k|k})$ using percentile curves, as well as the line corresponding to $\kappa(S) = 1.5028$. As can be clearly seen from this figure, the distribution of $\kappa(P_{k|k})$ quickly converges to the value of $\kappa(S) = 1.5028$, in accordance with the prediction of Eq. (100).

As an additional verification of Eq. (98), the following matrix relative discrepancy index is adopted:

$$J(A, B) \triangleq \frac{\sigma_{\max}(A - B)}{\sigma_{\max}(A)}, \qquad A, B \in \mathbb{R}^{n\times n} \tag{102}$$

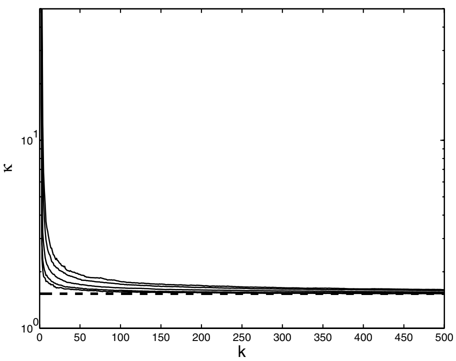


Fig. 5 Distribution (percentile curves) of the condition number of the covariance computed by the AEKF in 1000 Monte Carlo runs vs the condition number of the *S* matrix, computed via 50,000 samples (dashed line). Solid lines, top to bottom, correspond to 95, 85, 50, 15, and 5 percentiles.

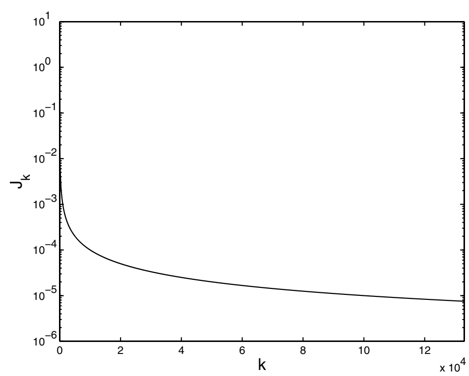


Fig. 6 Matrix discrepancy index; $J_k \triangleq J(S, (kP_{k|k})^{-1})$; single run.

Figure 6 shows the behavior of the discrepancy index $J(S, (kP_{k|k})^{-1})$ in a typical single run. Clearly, this index tends to zero, in agreement with the prediction of Eq. (98).

VI. Conclusions

The asymptotic properties of the quaternion estimation error covariance are investigated. It is proved that, although this covariance is regular, it does become asymptotically ill-conditioned as its trace decreases. In particular, it is shown that the asymptotic condition number of this covariance matrix is of the order of the inverse of its largest eigenvalue. A particularly interesting outcome of the analysis is that the aforementioned asymptotic behavior cannot be captured by low-order Taylor approximations of the covariance, such as the one computed by the additive extended Kalman filter. A geometric interpretation of these results is given using concepts from differential geometry.

The analytical results are backed by a simulation study employing both the extended Kalman filter and the recently introduced quaternion particle filter. In addition, the covariance computed by the additive extended Kalman filter is demonstrated to have a perfectly healthy condition number in a static, noise-free attitude estimation problem, contrary to predictions made in the literature.

Appendix A: Proof of Proposition 1

Proof. Let the indicator rv of an event C be defined as

$$\boldsymbol{I}_C(\omega) \triangleq \begin{cases} 1 & \omega \in C \\ 0 & \text{otherwise} \end{cases} \tag{A1}$$

Using the indicator function, the expectation of the random variable $(\boldsymbol{v}^T\boldsymbol{v})^{d+1}$ can be written as

$$E[(\boldsymbol{v}^T\boldsymbol{v})^{d+1}] = E\left[\sum_{i=1}^{\infty} (\boldsymbol{v}^T\boldsymbol{v})^{d+1} \boldsymbol{I}_{\{(i-1)\mathrm{tr}\,\Sigma < \boldsymbol{v}^T\boldsymbol{v} \le i\,\mathrm{tr}\,\Sigma\}}\right] \tag{A2}$$

Manipulating Eq. (A2) yields

$$\begin{aligned} E[(\boldsymbol{v}^T\boldsymbol{v})^{d+1}] &\le E\left[\sum_{i=1}^{\infty} i^{d+1} (\mathrm{tr}\,\Sigma)^{d+1} \boldsymbol{I}_{\{(i-1)\mathrm{tr}\,\Sigma < \boldsymbol{v}^T\boldsymbol{v} \le i\,\mathrm{tr}\,\Sigma\}}\right] \\ &= (\mathrm{tr}\,\Sigma)^{d+1}\left(\sum_{i=1}^{\infty} i^{d+1} E[\boldsymbol{I}_{\{(i-1)\mathrm{tr}\Sigma < \boldsymbol{v}^T\boldsymbol{v} \le i\,\mathrm{tr}\,\Sigma\}}]\right) \\ &= (\mathrm{tr}\,\Sigma)^{d+1}\left(\sum_{i=1}^{\infty} i^{d+1} \mathrm{prob}((i-1)\mathrm{tr}\,\Sigma < \boldsymbol{v}^T\boldsymbol{v} \le i\,\mathrm{tr}\,\Sigma)\right) \end{aligned} \tag{A3}$$

Because Σ is finite, it follows that $\mathrm{tr}\,\Sigma < \infty$. Under the condition stated in Proposition 1, Eq. (A3) implies the proposition. □

Appendix B: Proof of Proposition 2

Proof: The mth-order Taylor expansion of $\varphi(\boldsymbol{v})$ about the nominal realization $\boldsymbol{V}^*$ of $\boldsymbol{v}$ is given by

$$\hat{\varphi}_{[m]} = \sum_{i=0}^{m} \frac{1}{i!} E[((\boldsymbol{v} - \boldsymbol{V}^*)^T \nabla_{\boldsymbol{u}})^i f(\boldsymbol{u})|_{\boldsymbol{u}=\boldsymbol{V}^*}] \tag{B1}$$

Obviously,

$$\hat{\varphi}_{[m]} \le \sum_{i=0}^{m} \frac{1}{i!} E[(|\boldsymbol{v} - \boldsymbol{V}^*|^T |\nabla_{\boldsymbol{u}}|)^i f(\boldsymbol{u})|_{\boldsymbol{u}=\boldsymbol{V}^*}] \tag{B2}$$

where $|\boldsymbol{x}|$ denotes a vector for which the elements are $|x_j|$ ($j = 1, \ldots, n$). The notation $|\nabla_{\boldsymbol{u}}|$ indicates that the gradient is evaluated using the absolute values of its components; thus, its jth component is given by

$$(|\nabla_{\boldsymbol{u}}| f(\boldsymbol{u}))_j = \left|\frac{\partial f(\boldsymbol{u})}{\partial u_j}\right| \tag{B3}$$

Equation (B2) satisfies

$$\hat{\varphi}_{[m]} \le \sum_{i=0}^{m} c_{k,i} E[\|\boldsymbol{v} - \boldsymbol{V}^*\|_1^i] \tag{B4}$$

where $\|\cdot\|_1$ denotes the 1-norm, and the constant $c_{k,i}$ is defined as

$$c_{k,i} \triangleq \frac{1}{i!} \max_{j_1,\ldots,j_n} \left|\frac{\partial^i f(\boldsymbol{u})}{\partial u_{j_1} \partial u_{j_2} \cdots \partial u_{j_n}}\right|_{\boldsymbol{u}=\boldsymbol{V}^*} \tag{B5}$$

Recalling that every two distinct norms on $\mathbb{R}^n$ are equivalent, Eq. (B4) can be written using the Euclidean norm $\|\cdot\|$ as

$$\begin{aligned} \hat{\varphi}_{[m]} &\le \alpha \sum_{i=0}^{m} c_{k,i} E[\|\boldsymbol{v} - \boldsymbol{V}^*\|^i] \\ &= \alpha \sum_{i=0}^{m} c_{k,i} E[((\boldsymbol{v} - \boldsymbol{V}^*)^T (\boldsymbol{v} - \boldsymbol{V}^*))^{i/2}] \end{aligned} \tag{B6}$$

where α is a norm equivalence factor. Because $\boldsymbol{v} - \boldsymbol{V}^*$ satisfies Proposition 1 for any $d \ge 0$, Eq. (B6) implies the proposition. □

Appendix C: Auxiliary Results

Lemma 2. The following inequality holds for any θ, satisfying $0 \le \theta \le 1$:

$$((1-\theta)^{1/2} - 1)^2 \le \theta^2 \tag{C1}$$

Proof. Let f: $\mathbb{R} \longrightarrow \mathbb{R}$ be a scalar function defined as

$$f(a) = a^3 - 3a \tag{C2}$$

It can be verified that the function $f(a)$ is a nonincreasing function on the interval [0, 1] with a minimum at $a = +1$. Because $f(+1) = -2$, it follows that

$$a^3 - 3a + 2 \geq 0, \qquad a \in [0, 1] \tag{C3}$$

Multiplying both sides of Eq. (C3) by a and rearranging terms yields

$$a^4 - 2a^2 \geq a^2 - 2a \tag{C4}$$

whence

$$(1 - a^2)^2 \geq (a - 1)^2 \tag{C5}$$

Finally, setting $a = (1-\theta)^{1/2} \in [0, 1]$ in Eq. (C5) yields the Lemma. □

Lemma 3. Let $\Gamma \in \mathbb{R}^{n\times n}$ be an orthogonal matrix written as

$$\Gamma = [\Psi, \boldsymbol{\gamma}], \qquad \Psi \in \mathbb{R}^{n\times(n-1)}, \qquad \boldsymbol{\gamma} \in \mathbb{R}^n \tag{C6}$$

then the following inequality holds for every vector $\boldsymbol{v}$:

$$\|\Psi^T \boldsymbol{v}\| \leq \|\boldsymbol{v}\| \tag{C7}$$

where $\|\cdot\|$ denotes the Euclidean norm.

Proof. The definition of the L_2-induced matrix norm implies

$$\|\Psi^T \boldsymbol{v}\| \leq \sigma_{\max}(\Psi^T)\|\boldsymbol{v}\| \tag{C8}$$

Because Γ is orthogonal, it follows that

$$\Psi\Psi^T = I_{n\times n} - \boldsymbol{\gamma}\boldsymbol{\gamma}^T \tag{C9}$$

The eigenvalues of the rank-1 matrix $\boldsymbol{\gamma}\boldsymbol{\gamma}^T$ consist of $n-1$ zeros and a single nonzero eigenvalue that equals one. Therefore, Eq. (C9) can be written as

$$\Psi\Psi^T = U\left(I_{n\times n} - \begin{bmatrix} 0_{(n-1)\times(n-1)} & 0_{(n-1)\times 1} \\ 0_{1\times(n-1)} & 1 \end{bmatrix}\right)U^T \tag{C10}$$

where $U \in \mathbb{R}^{n\times n}$ is an orthogonal matrix having $\boldsymbol{\gamma}$ as its nth column. Equation (C10) implies that

$$\lambda_{\max}(\Psi\Psi^T) = 1 \tag{C11}$$

yielding $\sigma_{\max}(\Psi^T) = 1$. Using the last result in Eq. (C8) yields the Lemma. □

Lemma 4. Let $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$ be a random vector and its realization, respectively. Let also $G_1(\boldsymbol{\omega})$ and $G_2(\boldsymbol{\omega})$ be some arbitrary $n \times n$ finite random matrices, and let $Z(\boldsymbol{\omega})$ be an $n \times n$ random matrix for which the entries satisfy

$$Z_{ij}(\boldsymbol{\omega}) = \mathcal{O}(\epsilon^d) \quad \text{with probability 1}, \qquad \epsilon \to 0 \tag{C12}$$

Then the entries of the matrix

$$S \triangleq E[G_1(\boldsymbol{\omega})Z(\boldsymbol{\omega})G_2(\boldsymbol{\omega})] \tag{C13}$$

satisfy

$$S_{ij} = \mathcal{O}(\epsilon^d), \qquad \epsilon \to 0 \tag{C14}$$

Proof. Let $\boldsymbol{g}_1(\boldsymbol{\omega})$ and $\boldsymbol{g}_2(\boldsymbol{\omega})$ be two finite random vectors taking values in $\mathbb{R}^n$. Then

$$\begin{aligned} E[\boldsymbol{g}_1(\boldsymbol{\omega})^T Z(\boldsymbol{\omega})\boldsymbol{g}_2(\boldsymbol{\omega})] &= \int_{-\infty}^{+\infty} \boldsymbol{g}_1(\boldsymbol{\Omega})^T Z(\boldsymbol{\Omega})\boldsymbol{g}_2(\boldsymbol{\Omega}) p_{\boldsymbol{\omega}}(\boldsymbol{\Omega})\mathrm{d}\boldsymbol{\Omega} \\ &\leq [\max_{\boldsymbol{\Omega}} \boldsymbol{g}_1(\boldsymbol{\Omega})^T Z(\boldsymbol{\Omega})\boldsymbol{g}_2(\boldsymbol{\Omega})] \int_{-\infty}^{+\infty} p_{\boldsymbol{\omega}}(\boldsymbol{\Omega})\,\mathrm{d}\boldsymbol{\Omega} \\ &= [\max_{\boldsymbol{\Omega}} \boldsymbol{g}_1(\boldsymbol{\Omega})^T Z(\boldsymbol{\Omega})\boldsymbol{g}_2(\boldsymbol{\Omega})] \end{aligned} \tag{C15}$$

where $p_{\boldsymbol{\omega}}(\cdot)$ denotes the pdf of $\boldsymbol{\omega}$. Because $\boldsymbol{g}_1(\boldsymbol{\omega}), \boldsymbol{g}_2(\boldsymbol{\omega}) < \infty$ with probability 1, Eq. (C15) implies

$$E[\boldsymbol{g}_1(\boldsymbol{\omega})^T Z(\boldsymbol{\omega})\boldsymbol{g}_2(\boldsymbol{\omega})] = \mathcal{O}(\epsilon^d), \qquad \epsilon \to 0 \tag{C16}$$

Setting $\boldsymbol{g}_1(\boldsymbol{\omega}) = G_1(\boldsymbol{\omega})^T \boldsymbol{e}_i$ and $\boldsymbol{g}_2(\boldsymbol{\omega}) = G_2(\boldsymbol{\omega})\boldsymbol{e}_j$ where $\boldsymbol{e}_m \in \mathbb{R}^n$ denotes a unit-norm vector with 1 as its mth element, Eq. (C16) yields

$$S_{ij} = \boldsymbol{e}_i^T E[G_1(\boldsymbol{\omega})Z(\boldsymbol{\omega})G_2(\boldsymbol{\omega})]\boldsymbol{e}_j = \mathcal{O}(\epsilon^d), \qquad \epsilon \to 0 \tag{C17}$$

thereby completing the proof. □

Lemma 5. If the observations sequence $\{\boldsymbol{r}_i\}_{i=1}^k$ consists of at least two distinct noncolinear vectors, then the convergent sum

$$S = \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^k H(\boldsymbol{q}, \boldsymbol{r}_i)^T R^{-1} H(\boldsymbol{q}, \boldsymbol{r}_i) \tag{C18}$$

is nonsingular [i.e., $\lambda_{\min}(S) > 0$].

Proof. Through direct calculation it can be verified that the matrix $H(\boldsymbol{q}, \boldsymbol{r}_i)$ of Eq. (89) satisfies

$$H(\boldsymbol{q}, \boldsymbol{r}_i)\Xi(\boldsymbol{q})A(\boldsymbol{q})\boldsymbol{r}_i = \boldsymbol{0}_{3\times 1} \tag{C19a}$$

and

$$H(\boldsymbol{q}, \boldsymbol{r}_i)H(\boldsymbol{q}, \boldsymbol{r}_i)^T = 4(\boldsymbol{r}_i^T \boldsymbol{r}_i)I_{3\times 3} \tag{C19b}$$

where

$$\Xi(\boldsymbol{q}) \triangleq \begin{bmatrix} q_4 I_{3\times 3} + [\boldsymbol{\varrho}\times] \\ -\boldsymbol{\varrho}^T \end{bmatrix} \tag{C20}$$

Equations (C19) imply that $H(\boldsymbol{q}, \boldsymbol{r}_i)$ is full rank (i.e., of rank 3) with null space spanned by

$$\ker H(\boldsymbol{q}, \boldsymbol{r}_i) = \operatorname{span}(\boldsymbol{v}_i) \tag{C21}$$

where $\boldsymbol{v}_i \triangleq \Xi(\boldsymbol{q})A(\boldsymbol{q})\boldsymbol{r}_i$. Using the well-known property [7]

$$\Xi(\boldsymbol{q})^T \Xi(\boldsymbol{q}) = I_{3\times 3} \tag{C22}$$

yields, for any two vectors $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$,

$$\boldsymbol{v}_i^T \boldsymbol{v}_j = \boldsymbol{r}_i^T A(\boldsymbol{q})^T \Xi(\boldsymbol{q})^T \Xi(\boldsymbol{q})A(\boldsymbol{q})\boldsymbol{r}_j = \boldsymbol{r}_i^T \boldsymbol{r}_j \tag{C23}$$

which means that $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are colinear if and only if $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are colinear. Both Eqs. (C21) and (C23) assert that if $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are noncolinear, the kernels of $H(\boldsymbol{q}, \boldsymbol{r}_i)$ and $H(\boldsymbol{q}, \boldsymbol{r}_j)$ do not intersect. This in turn implies that as long as the sequence $\{\boldsymbol{r}_i\}_{i=1}^k$ consists of at least two distinct noncolinear observation vectors, the convergent sum in Eq. (C18) is nonsingular.

Appendix D: EKF Covariance Update

The EKF is a linear sequential estimator of the form

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + K_k(\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1}) \tag{D1}$$

where $K_k \in \mathbb{R}^{n\times p}$ and $H_k \in \mathbb{R}^{p\times n}$ are the KF-optimal gain matrix and measurement sensitivity matrix, respectively. The propagated and updated states at time $k-1$ are denoted by $\hat{\boldsymbol{x}}_{k|k-1}$ and $\hat{\boldsymbol{x}}_{k-1|k-1}$, respectively. Using Eq. (D1), the EKF estimation error is

$$\tilde{\boldsymbol{x}}_{k|k} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} - K_k(\boldsymbol{y}_k - H_k\hat{\boldsymbol{x}}_{k|k-1}) \tag{D2}$$

The KF assumes a linear measurement model of the form

$$\boldsymbol{y}_k = H_k\boldsymbol{x}_k + \boldsymbol{n}_k \tag{D3}$$

where $\boldsymbol{n}_k$ is zero-mean white Gaussian noise with a known covariance R_k. Its formal covariance measurement-update form is obtained by first substituting Eq. (D3) in Eq. (D2): that is,

$$\begin{aligned} \tilde{\boldsymbol{x}}_{k|k} &= (I - K_kH_k)\tilde{\boldsymbol{x}}_{k|k-1} - K_k\boldsymbol{n}_k \triangleq \mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1}), \\ \bar{\boldsymbol{x}}_k &= [\boldsymbol{x}_k^T, \boldsymbol{n}_k^T]^T \end{aligned} \tag{D4}$$

where $\tilde{\boldsymbol{x}}_{k|k-1} \triangleq \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}$ is the propagated estimation error at time $k-1$. Note that this rv is measurable on the joint probability space of x_k and $\mathcal{Y}^{k-1}$. Taking the conditional expectation of Eq. (D4) yields

$$\hat{\tilde{\boldsymbol{x}}}_{k|k} \triangleq E[\tilde{\boldsymbol{x}}_{k|k} | \mathcal{Y}^k = Y^k] = (I - K_k H_k) E[\tilde{\boldsymbol{x}}_{k|k-1} \mid \mathcal{Y}^{k-1} = Y^{k-1}] \tag{D5}$$

The EKF assumes that the conditional mean of the propagated estimation error is zero (an assumption that is essentially incorrect): that is,

$$E[\tilde{\boldsymbol{x}}_{k|k-1} \mid \mathcal{Y}^{k-1} = Y^{k-1}] = \mathbf{0}_{n\times 1} \tag{D6}$$

Under this assumption, Eq. (D5) yields

$$\hat{\tilde{\boldsymbol{x}}}_{k|k} = \mathbf{0}_{n\times 1} \tag{D7}$$

which implies that the computed estimation error covariance at time k is the second-moment matrix of the estimation error $\tilde{\boldsymbol{x}}_{k|k}$.

In general, the Taylor expansion of the second moment of $\tilde{\boldsymbol{x}}_{k|k}$ about $\bar{\boldsymbol{x}}_k^*$ takes the form

$$\begin{aligned} P_{k|k} &\triangleq E[\tilde{\boldsymbol{x}}_{k|k}\tilde{\boldsymbol{x}}_{k|k}^T | \mathcal{Y}^k = Y^k] = \mathcal{G}(\bar{\boldsymbol{x}}_k^*, Y^{k-1})\mathcal{G}(\bar{\boldsymbol{x}}_k^*, Y^{k-1})^T \\ &+ (\nabla_{\boldsymbol{x}_k^*}\mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1})^T)^T E[\boldsymbol{x}_k - \boldsymbol{x}_k^* | \mathcal{Y}^k = Y^k]\mathcal{G}(\bar{\boldsymbol{x}}_k^*, Y^{k-1})^T \\ &+ \mathcal{G}(\bar{\boldsymbol{x}}_k^*, Y^{k-1})(E[\boldsymbol{x}_k - \boldsymbol{x}_k^* | \mathcal{Y}^k = Y^k])^T \nabla_{\boldsymbol{x}_k^*}\mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1})^T \\ &+ (\nabla_{\boldsymbol{x}_k^*}\mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1})^T)^T E[(\boldsymbol{x}_k - \boldsymbol{x}_k^*)(\boldsymbol{x}_k - \boldsymbol{x}_k^*)^T | \mathcal{Y}^k = Y^k] \\ &\times \nabla_{\boldsymbol{x}_k^*}\mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1})^T + (\nabla_{\boldsymbol{n}_k^*}\mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1})^T)^T \\ &\times E[(\boldsymbol{n}_k - \boldsymbol{n}_k^*)(\boldsymbol{n}_k - \boldsymbol{n}_k^*)^T]\nabla_{\boldsymbol{n}_k^*}\mathcal{G}(\bar{\boldsymbol{x}}_k, \mathcal{Y}^{k-1})^T + \cdots \end{aligned} \tag{D8}$$

where $\nabla_{\boldsymbol{a}^*}\boldsymbol{f}^T$ is the Jacobian matrix with $\partial f_i/\partial a_j$ as its (i, j) element, evaluated at $\boldsymbol{a}^*$. Notice that in the EKF case, the third- and higher-order terms in Eq. (D8) vanish due to the linearity of the estimation error in Eq. (D4). In addition, because $\boldsymbol{n}_k$ is white, the cross-correlation terms vanish as well. The EKF sets the nominal state to $\bar{\boldsymbol{x}}_k^* = [\hat{\boldsymbol{x}}_{k|k-1}^T, \mathbf{0}_{1\times p}]^T$, which consequently means that the zeroth- and first-order terms in Eq. (D8) equal zero. This yields the conventional measurement-update form

$$P_{k|k} = (I - K_k H_k)P_{k|k-1}(I - K_k H_k)^T + K_k R K_k^T \tag{D9}$$

where the propagated estimation error covariance $P_{k|k-1}$ is defined as

$$P_{k|k-1} \triangleq E[\tilde{\boldsymbol{x}}_{k|k-1}\tilde{\boldsymbol{x}}_{k|k-1}^T \mid \mathcal{Y}^{k-1} = Y^{k-1}] \tag{D10}$$

Equation (D8) clearly shows that the covariance matrix of the EKF consists of second-order terms only.

Acknowledgments

This research was supported by the Israel Science Foundation under grant no. 1032/04 and by the Technion–Israel Institute of Technology's Asher Space Research Fund. The authors thank John L. Junkins for his comment that motivated Sec. IV and F. Landis Markley for helpful discussions.

References

[1] 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, Sept.–Oct. 1982, pp. 417–429. doi:10.2514/3.56190

[2] Crassidis, J. L., and Markley, F. L., "A Minimum Model Error Approach for Attitude Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 6, Nov.–Dec. 1997, pp. 1241–1247.

[3] Psiaki, M. L., "Attitude Determination Filtering Via Extended Quaternion Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 2, Mar.–Apr. 2000, pp. 210–214.

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

[5] Bak, T., Wisniewski, R., and Blanke, M., "Autonomous Attitude Determination and Control System for the Ørsted Satellite," *Proceedings of the IEEE Aerospace Applications Conference*, Inst. of Electrical and Electronics Engineers, Piscataway, NJ, Feb. 1996, pp. 173–186.

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

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

[8] Markley, F. L., and Sedlak, J. E., "Kalman Filter for Spinning Spacecraft Attitude Estimation," AIAA Guidance, Navigation, and Control Conference, Hilton Head, SC, AIAA Paper AIAA 2007-6812, Aug. 2007.

[9] Psiaki, M. L., "Estimation Using Quaternion Probability Densities on the Unit Hypersphere," *Malcolm D. Shuster Astronautics Symposium*, edited by J. L. Crassidis, F. L. Markley, J. L. Junkins, and K. C. Howell, Advances in the Astronautical Sciences, Vol. 122 Univelt, San Diego, CA, 2006; also American Astronautical Society Paper 05-461, 2005.

[10] Markley, F. L., "Attitude Filtering on SO(3)," *Malcolm D. Shuster Astronautics Symposium*, edited by J. L. Crassidis, F. L. Markley, J. L. Junkins, and K. C. Howell, Advances in the Astronautical Sciences, Vol. 122 Univelt, San Diego, CA, 2006; also American Astronautical Society Paper 05-460, 2005.

[11] Bar-Itzhack, I. Y., and Oshman, Y., "Attitude Determination from Vector Observations: Quaternion Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-21, No. 1, Jan. 1985, pp. 128–136. doi:10.1109/TAES.1985.310546

[12] Bar-Itzhack, I. Y., and Thienel, J. K., "On the Singularity in the Estimation of the Quaternion-of-Rotation," AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Monterey, CA, AIAA Paper 2002-4831, Aug. 2002.

[13] Calise, A. J., "Enforcing an Algebraic Constraint in Extended Kalman Filter Design," Proceedings of AIAA Guidance, Navigation, and Control Conference, Hilton Head, SC, AIAA, Paper 2007-6515, Aug. 2007.

[14] Zanetti, R., and Bishop, R. H., "Quaternion Estimation and Norm Constrained Kalman Filtering," Proceedings of AIAA Guidance Navigation and Control Conference, Keystone, CO, AIAA Paper 2006-6164, Aug. 2006.

[15] Oshman, Y., and Carmi, A., "Attitude Estimation from Vector Observations Using a Genetic Algorithm-Embedded Quaternion Particle Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 4, July–Aug. 2006, pp. 879–891. doi:10.2514/1.17951

[16] Doucet, A., Godsill, S., and Andrieu, C., "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, Vol. 10, No. 3, July 2000, pp. 197–208. doi:10.1023/A:1008935410038

[17] Estrada, R., and Kanwal, R. P., *Asymptotic Analysis: A Distributional Approach*, Birkhauser, Boston, 1994.

[18] Rudin, W., *Principles of Mathematical Analysis*, 3rd ed., McGraw–Hill, New York, 1976.

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

[20] Mendel, J. M., *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Signal Processing Series, Prentice–Hall, Upper Saddle River, NJ, 1995.