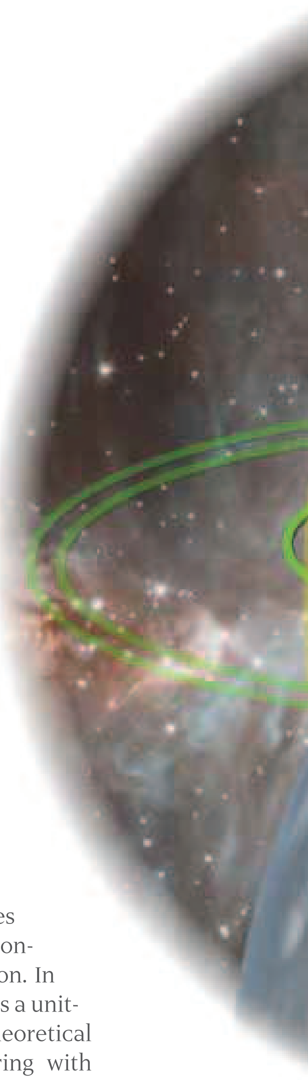
Norm-Constrained Kalman Filtering*

Renato Zanetti, Manoranjan Majji, Robert H. Bishop, and Daniele Mortari

Published in *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 5, September - October 2009, pp. 1458 - 1465.



ABSTRACT

The problem of estimating the state vector of a dynamical system from vector measurements when it is known that the state vector satisfies norm equality constraints is considered. The case of a linear dynamical system with linear measurements subject to a norm equality constraint is discussed with a review of existing solutions. The norm constraint introduces a nonlinearity in the system for which a new estimator structure is derived by minimizing a constrained cost function. It is shown that the constrained estimate is equivalent to the brute force normalization of the unconstrained estimate. The obtained solution is extended to nonlinear measurement models and applied to the spacecraft attitude filtering problems.

Introduction

It is well-known that the Kalman filter provides the *unconstrained* optimal solution of the linear stochastic estimation problem [1],[2]. The Kalman filter algorithm has two main phases: the state estimate propagation phase between measurements and the state estimate update phase when measurements become available. Unconstrained implies that the optimal state estimate is not constrained during the state estimate update phase as the measurements are processed. The Kalman filter provides the optimal state estimate considering n degrees of freedom (that is, the entire vector space $\Re^n$). However, if r state constraints are applied, the degrees of freedom are reduced to $n - r$. Projecting the unconstrained solution into the constrained space will not guarantee optimality. This work focuses on norm constraints applied to the state vector.

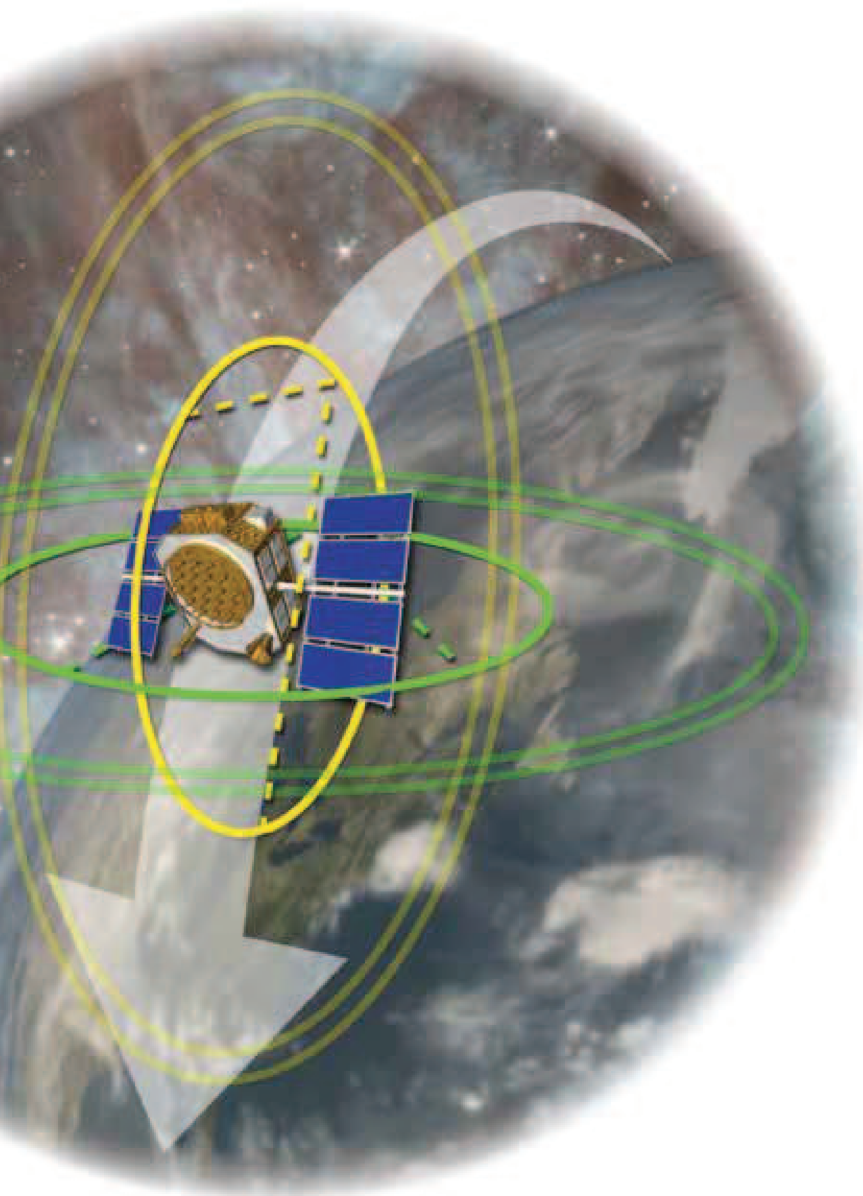
It is assumed throughout this paper that through the mathematical model (that is, through the state equation), the underlying physics, including the state constraints, are satisfied during periods between measurements. The mathematical model of the system should adequately represent any desired state constraints.

The objective of this work is to modify the Kalman filter solution to constrain the state update appropriately. In the discrete formulation of the Kalman filter, the state can be related to the control algebraically. The optimization problem is formulated as a parameter optimization problem; therefore, the state constraint can be expressed as a control constraint.

A motivation to seek the norm-constrained solution to the filtering problem is attitude estimation. Attitude estimation has been the topic of much research and debate in the past two decades [3]. The interest arises from the fact that the representation of the attitude is not a vector space and redundancy is necessary to avoid singularities and discontinuities [4]. For real-time space applications, the quaternion-of-rotation is the preferred attitude representation. In order to represent a rotation, the quaternion obeys a unit-norm constraint. This work will develop the theoretical foundations of norm-constrained Kalman filtering with reference to the quaternion estimation problem.

One method of introducing state constraints is to use pseudomeasurements [5]. The fundamental idea is to introduce a perfect measurement (hence the use of the term “pseudomeasurement”) consisting of the constraint equation into the estimation solution. This approach has shortcomings. The use of a perfect measurement results in a singular estimation problem known to occur when processing noise-free measurements in a Kalman filter. A

*Presented as paper AIAA 2006-6164 at the 2006 AIAA/AAS Astrodynamics Specialist Conference and paper AAS 08-215 at the 2008 AAS/AIAA Space Flight Mechanics Meeting.



small noise can be added to the pseudomeasurement to address the singularity; however, with the noise introduced, the constraint is no longer exactly satisfied.

One can consider state constraints when considering the optimization problems based on least squares methods. The solution to the least squares problem in the presence of linear equality constraints is found in Lawson and Hanson [6]. Another approach is to project the Kalman solution into the desired subspace. Since the projection can be done in different ways, a performance index can be defined to find the optimal projection. The optimal projection for the linear state equality constraint problem is presented in Simon and Chia [7]. The projection of the Kalman solution can be done at any time, not only during the update.

The constrained quaternion estimation problem was posed as a nonlinear programming problem by Psiaki [8] and solved using Newton's method. Psiaki's approach differs from the proposed approach in several key areas. First, Psiaki minimizes a different cost function and his method has a different interpretation of the covariance. Psiaki's method is a global optimal that solves a quadratically constrained quadratic program at every update stage that will work with poor or no initial estimate. The method proposed here is a prediction correction technique that relies on *a priori* estimates.

In sequential real-time quaternion estimation, two main approaches received the most attention: the Additive Extended Kalman Filter (AEKF) [9] and the Multiplicative Extended Kalman Filter (MEKF) [10]. Both the AEKF and MEKF necessitate restoring the norm constraint after the update. The straightforward method is to scale the updated quaternion by its norm, thereby minimizing the Euclidean distance between the unconstrained and the constrained estimates [11]. The main focus of this work is to obtain the optimal estimate while simultaneously constraining the norm. The result is that the normalization process provides the unitary estimate with minimum mean square error–a fact heretofore unproven. Previous work on the AEKF assumed quaternion normalization and studied the consequences [9],[12],[13]. In this work, normalization is not assumed, but is a direct result of the optimization process.

The paper is organized as follows. "Norm-Constrained Kalman Filtering" develops the new filter for a general norm constraint assuming linear dynamics and a linear measurement model. The problem is nonlinear because of the quadratic norm constraint. "Constrain Only Part of the State" shows how a subset of the state vector can be estimated using the norm-constrained algorithm. "Attitude Estimation" details the quaternion estimation problem used for the numerical examples. The results of "Norm-Constrained Kalman Filtering" are extended to a nonlinear measurement model for this example. "Numerical Results" contains numerical simulations of the new algorithm applied to quaternion estimation. "Conclusions" summarizes the work and develops some conclusions.

Norm-Constrained Kalman Filtering

Given a state that evolves through linear dynamics and given linear measurements, the optimal estimate in a mean-square error (MSE) sense is obtained using the Kalman filter algorithm. A norm-constrained estimate can be obtained by normalization of the Kalman (unconstrained) estimate. In this section, it will be shown that brute force normalization is *optimal* in an MSE sense. Normalization is a nonlinear transformation, therefore, similar approximations to those associated with the extended Kalman filter are made. Optimality does not hold strictly, but conditionally on the above approximations. The proof follows that of [14]. More recently, Julier and La Viola [15] presented a new method of Kalman filtering in the presence of nonlinear state variable equality constraints. They propose a method involving two projections (the first projection constrains the entire distribution and the second constrains the statistics) and they provide a comparison of the resulting estimates with other methods.

Define the *a priori* state estimate $\hat{\mathbf{x}}_k^-$ to be the state estimate at time t_k just prior to employing the measurement $\mathbf{y}_k$ in the state estimate update algorithm. The measurement model is given by

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \eta_k,$$

where η_k is measurement noise. Define the *a posteriori* state estimate $\hat{\mathbf{x}}_k^+$ to be the state estimate at time t_k just after the state estimate update. The performance index is defined as

$$\mathcal{J}_k = \mathrm{E}\{(\mathbf{e}_k^+)^{\mathrm{T}}\,\mathbf{e}_k^+\}, \tag{1}$$

which is the MSE of the estimator. The *a priori* and *a posteriori* estimation errors are given by

$$\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^- \quad \text{and} \quad \mathbf{e}_k^+ = \mathbf{x}_k - \hat{\mathbf{x}}_k^+,$$

respectively. Associated with the estimation errors, we can define the matrices

$$\mathbf{P}_k^- = \mathrm{E}\{\mathbf{e}_k^-\,(\mathbf{e}_k^-)^{\mathrm{T}}\}, \quad \text{and} \quad \mathbf{P}_k^+ = \mathrm{E}\{\mathbf{e}_k^+\,(\mathbf{e}_k^+)^{\mathrm{T}}\}, \tag{2}$$

before and after the measurement update, respectively. The matrices $\mathbf{P}_k^+$ and $\mathbf{P}_k^-$ are mean squares. For linear filters, when the mean of the estimation error is zero, these mean squares reduce to state estimate error covariances. However, in general for the nonlinear problem considered here, the estimator is not necessarily unbiased. The nonlinearity is introduced by the norm constraint of the state vector, which is desired to have a predefined value

$$\|\hat{\mathbf{x}}_k^+\| = \sqrt{l}$$

for the quaternion estimation problem $l = 1$. This constraint is equivalent to

$$(\hat{\mathbf{x}}_k^+)^{\mathrm{T}}\hat{\mathbf{x}}_k^+ = l \tag{3}$$

Notice that

$$\mathcal{J}_k = \mathrm{trace}\,\mathbf{P}_k^+.$$

The update is $\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\varepsilon_k$,

where $\varepsilon_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$ is the residual. The estimated measurement is given by

$$\hat{\mathbf{y}}_k = \mathbf{H}_k\hat{\mathbf{x}}_k.$$

Substituting the residual into Eq. (3), the state constraint can be expressed more conveniently as a control constraint:

$$\varepsilon_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\mathbf{K}_k\varepsilon_k + 2\hat{\mathbf{x}}_k^{-\mathrm{T}}\mathbf{K}_k\varepsilon_k + \hat{\mathbf{x}}_k^{-\mathrm{T}}\hat{\mathbf{x}}_k^- - l = 0. \tag{4}$$

The goal is to find the gain $\mathbf{K}_k$ such that Eq. (1) is minimized and the constraint given by Eq. (4) is satisfied.

First-Order Condition

The *a posteriori* error mean square is given by the Joseph formula:*

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}}$$

where $\mathbf{P}_k^-$ is the *a priori* state error mean square and $\mathbf{R}_k$ is the covariance of η_k, assumed to be zero mean. Define

$$\mathbf{W}_k \triangleq \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k.$$

The Joseph formula can be rewritten as

$$\mathbf{P}_k^+ = \mathbf{P}_k^- - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^- - \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{W}_k\mathbf{K}_k^{\mathrm{T}}.$$

The performance index to be minimized is then given by

$$\mathcal{J}_k = \mathrm{trace}\,[\mathbf{P}_k^- - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^- - \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{W}_k\mathbf{K}_k^{\mathrm{T}}].$$

The Kalman gain is computed to satisfy the constraint in Eq. (4). We have $\mathbf{P}_k^- \in \Re^{n\times n}$, $\mathbf{K}_k \in \Re^{n\times m}$, $l \in \Re$, and the remaining are of appropriate dimensions. The augmented performance index is

$$\mathcal{J}_k = \mathrm{trace}\,[\mathbf{P}_k^- - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^- - \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{W}_k\mathbf{K}_k^{\mathrm{T}}] + \lambda_k\,[(\hat{\mathbf{x}}_k^-)^{\mathrm{T}}\hat{\mathbf{x}}_k^- + 2\varepsilon_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\hat{\mathbf{x}}_k^- + \varepsilon_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\mathbf{K}_k\varepsilon_k - l\,].$$

The $n \times m + 1$ optimal values of λ_k and $\mathbf{K}_k$ are obtained solving the $n \times m$ equations resulting from taking the derivative of $\mathcal{J}_k$ with respect to $\mathbf{K}_k$ and setting it to zero yielding

$$-2\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + 2\mathbf{K}_k\mathbf{W}_k + 2\lambda_k(\hat{\mathbf{x}}_k^-\varepsilon_k^{\mathrm{T}} + \mathbf{K}_k\varepsilon_k\varepsilon_k^{\mathrm{T}}) = 0, \tag{5}$$

and the scalar constraint Eq. (4). The vector and matrix derivatives used to obtain Eq. (5) are listed in the appendix.

Eq. (5) can be rewritten to obtain the following first-order conditions

$$\mathbf{K}_k = (\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \lambda_k\hat{\mathbf{x}}_k^-\varepsilon_k^{\mathrm{T}})(\mathbf{W}_k + \lambda_k\varepsilon_k\varepsilon_k^{\mathrm{T}})^{-1},$$

and

$$\varepsilon_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\mathbf{K}_k\varepsilon_k + 2(\hat{\mathbf{x}}_k^-)^{\mathrm{T}}\,\mathbf{K}_k\varepsilon_k + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}}(\hat{\mathbf{x}}_k^-) - l = 0.$$

Using the matrix inversion lemma (see appendix), it follows that

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\mathbf{W}_k^{-1} - \lambda_k\hat{\mathbf{x}}_k^-\varepsilon_k^{\mathrm{T}}\mathbf{W}_k^{-1} - \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\frac{\lambda_k\varepsilon_k\varepsilon_k^{\mathrm{T}}\mathbf{W}_k^{-1}}{1 + \lambda_k\varepsilon_k^{\mathrm{T}}\mathbf{W}_k^{-1}\varepsilon_k} + \lambda_k\hat{\mathbf{x}}_k^-\varepsilon_k^{\mathrm{T}}\mathbf{W}_k^{-1}\frac{\lambda_k\varepsilon_k\varepsilon_k^{\mathrm{T}}\mathbf{W}_k^{-1}}{1 + \lambda_k\varepsilon_k^{\mathrm{T}}\mathbf{W}_k^{-1}\varepsilon_k}.$$

*see the text immediately following Eq. (10) for a discussion of the validity of the Joseph formula for this particular problem.

Substituting into Eq. (4), after some manipulations, the following scalar equation with the scalar unknown λ_k is obtained

$$\lambda_k^2 \tilde{\varepsilon}_k^2 \left(-(\hat{\mathbf{x}}_k^-)^{\mathrm{T}} \hat{\mathbf{x}}_k^- + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}} (\hat{\mathbf{x}}_k^-) - l\right) + \lambda_k \tilde{\varepsilon}_k \left(-2\,(\hat{\mathbf{x}}_k^-)^{\mathrm{T}} \hat{\mathbf{x}}_k^- \right.$$
$$\left. + 2\,(\hat{\mathbf{x}}_k^-)^{\mathrm{T}} (\hat{\mathbf{x}}_k^-) - 2l\right) + \left(\varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \mathbf{H}_k \mathbf{P}_k^- \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k \right.$$
$$\left. + 2\,(\hat{\mathbf{x}}_k^-)^{\mathrm{T}} \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}} (\hat{\mathbf{x}}_k^-) - l\right) = 0\,,$$

where

$$\tilde{\varepsilon}_k = \varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k\,.$$

Therefore, the optimal Lagrange multiplier is

$$\lambda_k = \frac{-b/2 \pm \sqrt{b^2/4 - ac}}{a}\,,$$

where

$$a = -l\tilde{\varepsilon}_k^2, \qquad b = -2\tilde{\varepsilon}_k l, \text{ and}$$

$$c = \varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \mathbf{H}_k \mathbf{P}_k^- \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k + 2\,(\hat{\mathbf{x}}_k^-)^{\mathrm{T}} \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}} (\hat{\mathbf{x}}_k^-) - l.$$

Finally, it follows that

$$\lambda_k = \frac{\tilde{\varepsilon}_k l \pm \sqrt{\tilde{\varepsilon}_k^2 l^2 + l\tilde{\varepsilon}_k^2 c}}{-l\tilde{\varepsilon}_k^2} = \frac{1 \pm \sqrt{1 + c/l}}{-\tilde{\varepsilon}_k}\,. \tag{6}$$

Notice that

$$1 + c/l = (\varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \mathbf{H}_k \mathbf{P}_k^- \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k + 2\,(\hat{\mathbf{x}}_k^-)^{\mathrm{T}} \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}}\ (\hat{\mathbf{x}}_k^-))\,/\,l$$
$$= (\varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \mathbf{H}_k \mathbf{P}_k^- + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}})^{\mathrm{T}} (\varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \mathbf{H}_k \mathbf{P}_k^- + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}})/l \geq 0$$

Therefore, λ_k is always a real number and can be rewritten as

$$\lambda_k = \frac{-1}{\tilde{\varepsilon}_k} \pm \frac{\|\hat{\mathbf{x}}_k^- + \mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} \mathbf{W}_k^{-1} \varepsilon_k\|}{\tilde{\varepsilon}_k \sqrt{l}}\,.$$

Second-Order Condition

Taking the second derivative of the performance index presents some representation issues. Each entry of the first derivative can be differentiated again, but this approach results in $m \times n$ matrix equations. Another approach is to perturb the gain and show that the perturbation results in an increment of the performance index. We will show the proof in the case of scalar measurement; the result is identical for vector measurements.

In the case of scalar measurement, the gain $\mathbf{K}_k$ reduces to a vector and can be partitioned as

$$\mathbf{K}_k = \begin{bmatrix} \mathbf{k}_k \\ k_k \end{bmatrix},$$

where k_k is a scalar. The constraint becomes

$$\varepsilon_k^2 \mathbf{k}_k^{\mathrm{T}} \mathbf{k}_k + \varepsilon_k^2\, k_k^2 + 2\varepsilon_k\,(\hat{X}_k^-)^{\mathrm{T}}\, \mathbf{k}_k + 2\varepsilon_k x_k^- k_k + (\hat{X}_k^-)^{\mathrm{T}}\, \hat{X}_k^- - l = 0,$$

where

$$\hat{\mathbf{x}}_k^- = \begin{bmatrix} \hat{X}_k^- \\ \hat{x}_k^- \end{bmatrix}.$$

Differentiating the constraint yields

$$2(\varepsilon_k^2 \mathbf{k}_k^{\mathrm{T}} + \varepsilon_k (\hat{X}_k^-)^{\mathrm{T}}) d\mathbf{k}_k + 2(\varepsilon_k^2 k_k + \varepsilon_k \hat{x}_k^-) dk_k = 0\,.$$

Assuming the residual is not zero (if the residual is zero, the *a posteriori* estimate is always equal to the *a priori* estimate), it follows that

$$dk_k = -\frac{\varepsilon_k \mathbf{k}_k^{\mathrm{T}} + (\hat{X}_k^-)^{\mathrm{T}}}{\varepsilon_k k_k + \hat{x}_k^-}\, d\mathbf{k}_k\,.$$

The second-order differential of the performance index is

$$d\mathcal{J}_k^2 = d\mathbf{K}_k^{\mathrm{T}} G_{KK} d\mathbf{K}_k\,, \tag{7}$$
$$G_{KK} = 2W_k + 2\lambda_k \varepsilon_k^2\,, \text{ and}$$
$$d\mathcal{J}_k^2 = (2W_k + 2\lambda_k \varepsilon_k^2)(d\mathbf{K}_k^{\mathrm{T}} d\mathbf{K}_k) = (2W_k + 2\lambda_k \varepsilon_k^2)(d\mathbf{k}_k^{\mathrm{T}} d\mathbf{k}_k + dk_n^2)$$
$$= d\mathbf{k}_k^{\mathrm{T}} \left\{ 2(W_k + \lambda_k \varepsilon_k^2) \left(\mathbf{I} + \frac{\varepsilon_k \mathbf{k}_k + \hat{X}_k^-}{\varepsilon_k k_k + \hat{x}_k^-} \frac{\varepsilon_k \mathbf{k}_k^{\mathrm{T}} + (\hat{X}_k^-)^{\mathrm{T}}}{\varepsilon k_k + \hat{x}_k^-} \right) \right\} d\mathbf{k}_k.$$

The sufficient condition for a minimum is that the matrix inside curly brackets in Eq. (7) is positive definite. An equivalent condition is

$$(W_k + \lambda_k \varepsilon_k^2) \left\{ \mathbf{I} + \left[\frac{\varepsilon_k + (\tilde{\mathbf{P}}_k^-)^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} + W_k \hat{X}_k^-}{(\varepsilon_k k_k + \hat{x}_k^-)(W_k + \lambda_k \varepsilon_k^2)} \right] \left[\frac{\varepsilon_k + (\tilde{\mathbf{P}}_k^-)^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} + W_k \hat{X}_k^-}{(\varepsilon_k k_k + \hat{x}_k^-)(W_k + \lambda_k \varepsilon_k^2)} \right]^{\mathrm{T}} \right\} > 0 \tag{8}$$

since

$$\mathbf{k}_k = \frac{(\tilde{\mathbf{P}}_k^-)^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} - \lambda_k \varepsilon_k \hat{X}_k^-}{W_k + \lambda_k \varepsilon_k^2},\ k_k = \frac{\mathbf{p}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} - \lambda_k \varepsilon_k \hat{x}_k^-}{W_k + \lambda_k \varepsilon_k^2}, \mathbf{P}_k^- = [\,\tilde{\mathbf{P}}_k^-\, \mathbf{p}_k\,].$$

The matrix in the curly brackets in Eq. (8) is of the form

$$\mathbf{I} + \mathbf{v}\mathbf{v}^{\mathrm{T}},$$

which is always positive definite. As a consequence, the optimal gain produces a minimum performance index when the scalar $W_k + \lambda_k \varepsilon_k^2$ is positive. Since

$$W_k + \lambda_k \varepsilon_k^2 = \pm\, W_k \sqrt{1 + c/l}\,,$$

the minimum occurs when the plus sign is chosen for the lagrange multiplier. Also, if the minus sign is chosen, the performance index will be maximized.

Constrained Minimum Solution

The performance index is minimized and the constraint is satisfied when the optimal gain is chosen as

$$\mathbf{K}_k^* = (\mathbf{P}_k^- \mathbf{H}_k^{\mathrm{T}} - \lambda_k \hat{\mathbf{x}}_k^- \varepsilon_k^{\mathrm{T}})(\mathbf{W}_k + \lambda_k \varepsilon_k \varepsilon_k^{\mathrm{T}})^{-1}\,,$$

where

$$\lambda_k = \frac{-1}{\tilde{\varepsilon}_k} \pm \frac{\|\,\varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1} \mathbf{H}_k \mathbf{P}_k^- + (\hat{\mathbf{x}}_k^-)^{\mathrm{T}}\|}{\tilde{\varepsilon}_k \sqrt{l}}.$$

The asterisk was added to distinguish from the unconstrained Kalman gain $\mathbf{K}_k$ given by

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k \mathbf{W}_k^{-1}\,.$$

The unconstrained *a posteriori* estimate is

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \varepsilon_k.$$

The minimizing constrained gain can be rewritten as

$$\mathbf{K}_k^* = \mathbf{K}_k + \left(\frac{\sqrt{l}}{\|\hat{\mathbf{x}}_k^+\|} - 1 \right) \hat{\mathbf{x}}_k^+ \frac{\varepsilon_k^{\mathrm{T}} \mathbf{W}_k^{-1}}{\tilde{\varepsilon}_k} \tag{9}$$

Property 1. *The optimal constrained solution shares the same direction as the optimal unconstrained solution.*

Proof. Let $\hat{\mathbf{x}}_k^*$ be the optimal constrained estimate. Then it follows that

$$\hat{\mathbf{x}}_k^* = \hat{\mathbf{x}}_k^- + \mathbf{K}_k^*\boldsymbol{\varepsilon}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\boldsymbol{\varepsilon}_k + \left(\frac{\sqrt{l}}{\|\hat{\mathbf{x}}_k^+\|} - 1\right)\hat{\mathbf{x}}_k^+ \frac{\boldsymbol{\varepsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}}{\tilde{\varepsilon}_k}\boldsymbol{\varepsilon}_k = \frac{\sqrt{l}}{\|\hat{\mathbf{x}}_k^+\|}\hat{\mathbf{x}}_k^+.$$

So $\hat{\mathbf{x}}_k^*$ and $\hat{\mathbf{x}}_k^+$ have the same direction, but different magnitude. Property 1 states that brute force normalization is optimal not only in a geometrical sense, but also in an MSE sense.

The *a posteriori* estimation error is

$$\mathbf{e}_k^* = (\mathbf{I} - \mathbf{K}_k^*\mathbf{H}_k)\mathbf{e}_k^- + \mathbf{K}_k^*\boldsymbol{\eta}_k.$$

Under the assumption that measurement noise $\boldsymbol{\eta}_k$ is independent of process noise and initial estimation error, it follows that

$$\mathbf{P}_k^* = \mathrm{E}\{(\mathbf{I} - \mathbf{K}_k^*\mathbf{H}_k)\mathbf{e}_k^-(\mathbf{e}_k^-)^{\mathrm{T}}(\mathbf{I} - \mathbf{K}_k^*\mathbf{H}_k)^{\mathrm{T}}\} + \mathrm{E}\{\mathbf{K}_k^*\boldsymbol{\eta}_k\boldsymbol{\eta}_k^{\mathrm{T}}(\mathbf{K}_k^*)^{\mathrm{T}}\}. \quad (10)$$

The optimal gain is a function of the *a priori* state and the residual, therefore it is a random variable and it should not be taken out the expectation operator. A similar situation happens in nonlinear Kalman filtering. In the extended Kalman filter, for example, the measurement mapping matrix is a function of the *a priori* state, thus making the gain a function of the *a priori* state as well. The straightforward approach is to take the Kalman gain out of the expectation sign (following the EKF derivation assumption) yielding

$$\mathbf{P}_k^* = (\mathbf{I} - \mathbf{K}_k^*\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k^*\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k^*\mathbf{R}_k(\mathbf{K}_k^*)^{\mathrm{T}}.$$

Therefore, $\mathbf{P}_k^*$ is a covariance-like matrix and not strictly the covariance of the estimation error. Substituting for $\mathbf{K}_k^*$ yields

$$\mathbf{P}_k^* = \mathbf{P}_k^+ + \frac{1}{\tilde{\varepsilon}_k}\left(1 - \frac{\sqrt{l}}{\|\hat{\mathbf{x}}_k^+\|}\right)^2 \hat{\mathbf{x}}_k^+(\hat{\mathbf{x}}_k^+)^{\mathrm{T}}, \quad (11)$$

which is very similar to the correction given by Choukroun *et al.* [16] The covariance update by Choukroun *et al.* is given by

$$\mathbf{P}_k^* = \mathbf{P}_k^+ + \left(1 - \frac{\sqrt{l}}{\|\hat{\mathbf{x}}_k^+\|}\right)^2 \hat{\mathbf{x}}_k^+(\hat{\mathbf{x}}_k^+)^{\mathrm{T}}.$$

Besides the evident difference of missing $\tilde{\varepsilon}_k$, the two methods also differ from a more fundamental perspective. Choukroun *et al.* assume brute force normalization and derive a covariance correction. In this work, brute force normalization is shown to be optimal in the mean square sense.

The proposed covariance correction is a second-order effect

$$\mathbf{P}_k^* = \mathbf{P}_k^+ + \frac{1}{\tilde{\varepsilon}_k}\frac{\hat{\mathbf{x}}_k^+(\hat{\mathbf{x}}_k^+)^{\mathrm{T}}}{\|\hat{\mathbf{x}}_k^+\|}(\|\hat{\mathbf{x}}_k^+\|) - \sqrt{l})^2.$$

Both the AEKF and MEKF provide estimates with unit norm to first order [17], since the EKF covariance is valid to first order, the correction is not necessary, and $\mathbf{P}_k^+$ can be used. If the AEKF necessitates a covariance adjustment, so does the first-order MEKF since brute force normalization affects the multiplicative error as well.

When two random variables are related through a nonlinear transformation, it is generally impossible to relate exclusively their second moments, but all the moments of the original variable will contribute in the second moment of the transformed variable. Therefore, the correction on the covariance can be accurate or not depending on the distribution of the estimation error. The matrix $\mathbf{P}_k^*$ is an approximation and, like any approximation, might not be satisfactory under certain circumstances. From Eq. (11) it can be seen that $\mathbf{P}_k^*$ can be unsatisfactory for small $\tilde{\varepsilon}_k$ and large norm errors of the unconstrained estimate. This situation could arise, for example, in the presence of scalar measurement when the estimation error is large.

Constrain Only Part of the State

The derivation of the previous section assumes the entire state $\mathbf{x}$ is subject to the norm constraint. In this section, it is shown how to constrain only part of the state. Suppose that the $n \times 1$ state vector $\mathbf{x}$ is partitioned into $\mathbf{z}$ and $\mathbf{q}$ as

$$\mathbf{x} = \begin{bmatrix}\mathbf{z}\\ \mathbf{q}\end{bmatrix},$$

where $\mathbf{z} \in \Re^{n-p}$ and $\mathbf{q} \in \Re^{p}$. Suppose that $\mathbf{z}$ is not subject to any constraint, while $\mathbf{q}$ needs to satisfy a norm constraint. The estimation error associated with each partition will be minimized independently. The Kalman gain is partitioned appropriately as

$$\mathbf{K} = \begin{bmatrix}\mathbf{K}_z\\ \mathbf{K}_q\end{bmatrix},$$

where $\mathbf{K}_z \in \Re^{(n-p)\times m}$, $\mathbf{K}_q \in \Re p \times m$, also m is the dimension of the measurement vector $\mathbf{y}$. At the measurement time t_k, a linear update is assumed where

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \begin{bmatrix}\mathbf{K}_{z,k}\\ \mathbf{K}_{q,k}\end{bmatrix}(\mathbf{y}_k - \hat{\mathbf{y}}_k).$$

The measurement model is

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\eta}_k = \mathbf{H}_k(\hat{\mathbf{x}}_k^- + \mathbf{e}_k^-) + \boldsymbol{\eta}_k = \hat{\mathbf{y}}_k + \mathbf{H}_k\mathbf{e}_k^- + \boldsymbol{\eta}_k.$$

The estimation error covariance before the update can be partitioned as follows

$$\mathbf{P}_k^- = [\mathbf{P}_{1,k} \;\; \mathbf{P}_{2,k}] = \begin{bmatrix}\mathbf{P}_{zz,k}^- & \mathbf{P}_{zq,k}^-\\ \mathbf{P}_{qz,k}^- & \mathbf{P}_{qq,k}^-\end{bmatrix}; \mathbf{P}_{2,k} \in \Re^{n\times p}.$$

Note that

$$\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^- = \begin{bmatrix}\mathbf{K}_{z,k}\mathbf{H}_k\mathbf{P}_{1,k} & \mathbf{K}_{z,k}\mathbf{H}_k\mathbf{P}_{2,k}\\ \mathbf{K}_{q,k}\mathbf{H}_k\mathbf{P}_{1,k} & \mathbf{K}_{q,k}\mathbf{H}_k\mathbf{P}_{2,k}\end{bmatrix}$$

$$\mathbf{K}\mathbf{R}_k\mathbf{K}^{\mathrm{T}} = \begin{bmatrix}\mathbf{K}_{z,k}\mathbf{R}_k\mathbf{K}_{z,k}^{\mathrm{T}} & \mathbf{K}_{z,k}\mathbf{R}_k\mathbf{K}_{q,k}^{\mathrm{T}}\\ \mathbf{K}_{q,k}\mathbf{R}_k\mathbf{K}_{z,k}^{\mathrm{T}} & \mathbf{K}_{q,k}\mathbf{R}_k\mathbf{K}_{q,k}^{\mathrm{T}}\end{bmatrix}.$$

The partitioned *a posteriori* covariance is

$$\begin{aligned}
\mathbf{P}_{zz,k}^+ &= \mathbf{P}_{zz,k}^- - \mathbf{K}_{z,k}\mathbf{H}_k\mathbf{P}_{1,k} - \mathbf{P}_{1,k}^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}K_{z,k}^{\mathrm{T}} + \mathbf{K}_{z,k}\mathbf{W}_k\mathbf{K}_{z,k}^{\mathrm{T}}\\
\mathbf{P}_{zq,k}^+ &= \mathbf{P}_{zq,k}^- - \mathbf{K}_{z,k}\mathbf{H}_k\mathbf{P}_{2,k} - \mathbf{P}_{1,k}^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_{q,k}^{\mathrm{T}} + \mathbf{K}_{z,k}\mathbf{W}_k\mathbf{K}_{q,k}^{\mathrm{T}}\\
\mathbf{P}_{qq,k}^+ &= \mathbf{P}_{qq,k}^- - \mathbf{K}_{q,k}\mathbf{H}_k\mathbf{P}_{2,k} - \mathbf{P}_{2,k}^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_{q,k}^{\mathrm{T}} + \mathbf{K}_{q,k}\mathbf{W}_k\mathbf{K}_{q,k}^{\mathrm{T}}\\
\mathbf{W}_k &= \mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k
\end{aligned}$$

The matrix $\mathbf{P}^+_{zz,k}$ is only a function of $\mathbf{K}_{z,k}$, and $\mathbf{P}^+_{qq,k}$ is only a function of $\mathbf{K}_{q,k}$. Also, the trace of $\mathbf{P}^+_k$ is equal to the sum of the traces of $\mathbf{P}^+_{zz,k}$ and $\mathbf{P}^+_{qq,k}$. The two facts imply that the minimum of the sum is equal to the sum of the minima, or

$$\begin{aligned}\min_{\mathbf{K}_k}(\text{trace } \mathbf{P}^+) &= \min_{\mathbf{K}_{z,k},\mathbf{K}_{q,k}}(\text{trace } \mathbf{P}^+_{zz,k} + \text{trace } \mathbf{P}^+_{qq,k}) \\ &= \min_{\mathbf{K}_{z,k}}(\text{trace } \mathbf{P}^+_{zz,k}) + \min_{\mathbf{K}_{q,k}}(\text{trace } \mathbf{P}^+_{qq,k}),\end{aligned}$$

hence, the two minimizations can be performed independently. The optimal gains are

$$\mathbf{K}_{z,k} = \mathbf{P}^T_{1,k}\mathbf{H}^T_k\mathbf{W}^{-1}_k \quad \text{and} \quad \mathbf{K}_{q,k} = \mathbf{P}^T_{2,k}\mathbf{H}^T_k\mathbf{W}^{-1}_k. \tag{12}$$

As expected, there is no difference in calculating the gains independently or together, because the correlation is taken into account in $\mathbf{P}_{1,k}$ and $\mathbf{P}_{2,k}$.

$$\mathbf{K}_k = \begin{bmatrix}\mathbf{K}_{z,k} \\ \mathbf{K}_{q,k}\end{bmatrix} = \begin{bmatrix}\mathbf{P}^T_{1,k} \\ \mathbf{P}^T_{2,k}\end{bmatrix}\mathbf{H}^T_k\mathbf{W}^{-1}_k = \mathbf{P}^{-1}_k\mathbf{H}^T_k\mathbf{W}^{-1}_k.$$

Generally, there is no advantage in computing the gain via the partition because the full residual covariance matrix still has to be inverted. However, when the updates of $\mathbf{q}$ and $\mathbf{z}$ are different, the optimal gain can be derived via the partitions and the results previously presented apply to calculate $\mathbf{K}_{q,k}$ to obtain a norm-constrained estimate of $\mathbf{q}$.

Attitude Estimation

An important application of the norm-constrained filter architecture is spacecraft attitude estimation. Mathematical models describing rotational motion of spacecraft typically constitute equations involving the kinematics and dynamics. Several alternative sets of parameters are available to model the rotational kinematics of a rigid body [18]. All of the three parameter sets to describe orientation possess a singularity. The quaternion of rotation, having four parameters, forms a singularity-free attitude parameterization of the lowest dimension. However, being once redundant, it is subject to unit norm constraint. This description of attitude kinematics and dynamics has lead to many successful spacecraft attitude estimation and control designs [3]. The equations governing the quaternion kinematics are given by

$$\dot{\mathbf{q}} = 1/2\,\Omega(\omega)\mathbf{q} \tag{13}$$

where $\mathbf{q}^T = \{\rho^T, q_4\} = \{q_1, q_2, q_3, q_4\}$ is the quaternion, $\omega^T = \{\omega_1, \omega_2, \omega_3\}$ is the angular velocity (body frame)

$$\Omega(\omega) = \begin{bmatrix}-[\omega\times] & \omega \\ -\omega^T & 0\end{bmatrix}, \quad \text{and} \quad [\omega\times] = \begin{bmatrix}0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0\end{bmatrix}.$$

As long as the direction of ω does not change, Eq. (13) admits closed-form solution [19] that can be used to propagate the quaternion from t_k to t_{k+1}

$$\mathbf{q}^-_{k+1} = \left[\mathbf{I}_{4\times4}\cos\left(\frac{\|\omega_k\|\,\delta t_k}{2}\right) + \frac{\Omega(\omega_k)}{\|\omega^+_k\|}\sin\left(\frac{\|\omega_k\|\,\delta t_k}{2}\right)\right]\mathbf{q}^+_k. \tag{14}$$

This equation is exact for a pure spin (about any axis). This condition is a good approximation in most practical cases.

Since the quaternion must preserve its length, Eq. (14) can be obtained by building the matrix performing a rigid rotation in a 4D space [20]. In fact, the quaternion describing the attitude of a pure-spinning rigid body also spins in the 4D space by describing a great circle with angular velocity $\omega/2$ [21]. The plane of rotation is defined by the 4×2 matrix $\mathbf{P} = [\mathbf{q} \;\; \dot{\mathbf{q}}/\|\dot{\mathbf{q}}\|]$. Equation (14) is obtained from the quaternion spin rate and the plane of rotation.*

The angular velocity, ω is available for measurement aboard spacecraft by gyros. In addition, there are attitude sensors (e.g., star trackers, magnetometers, Sun sensors, etc.) providing vector. The rate gyro measurement, typically modeled as a random walk following the classical work of Farrenkopf [23], is given by

$$\begin{aligned}\tilde{\omega} &= \omega + \beta + \eta_v \\ \dot{\beta} &= \eta_u\end{aligned}$$

where η_v and η_u are zero mean white noise processes with covariances given by

$$E\{\eta_v(t)\,\eta^T_v(\tau)\} = \sigma^2_v\mathbf{I}_{3\times3}\delta(t-\tau) \;\text{ and }\; E\{\eta_u(t)\,\eta^T_u(\tau)\} = \sigma^2_u\mathbf{I}_{3\times3}\delta(t-\tau),$$

respectively, and β represents the gyro bias and random walk. The attitude sensor measurement model is given by

$$\mathbf{b}_i = \mathbf{A}(\mathbf{q})\mathbf{r}_i + \mathbf{v}_i,$$

where $\mathbf{b}_i$ and $\mathbf{r}_i$ are the i^{th} body vector and the reference vector, respectively, and $\mathbf{A}(\mathbf{q})$ is the direction cosine matrix parameterized in terms of the unknown true quaternion $\mathbf{q}$.

The attitude estimation problem is nonlinear (in general) owing to the presence of nonlinearity in the quaternion kinematics and measurement models. In this section, the novel filtering architecture (termed constrained Kalman filter (CKF) for the subsequent developments of the paper) discussed in the previous sections is applied to attitude filtering. The developments of the previous sections assume a linear dynamical system and a linear measurement model in the presence of a quadratic state equality constraint. The results obtained for the linear case are extended to nonlinear dynamics and measurements through the use of linearization, as in the extended Kalman filter.

To this effect, consider the quaternion estimation error given by the quaternion product

$$\delta\mathbf{q} = \mathbf{q} \otimes \hat{\mathbf{q}}^{-1}. \tag{15}$$

where $\delta\mathbf{q}^T = \{\delta\rho^T, \delta q_4\}$. Using the properties of the quaternion product and quaternion kinematics, it can be shown that [10]

$$\delta\dot{\mathbf{q}} = \begin{bmatrix}[\hat{\omega}\times] & \delta\rho \\ & 0\end{bmatrix} + 1/2\begin{bmatrix}\delta\omega \\ 0\end{bmatrix} \otimes \delta\mathbf{q}. \tag{16}$$

The first-order approximation of Eq. (16)

$$\begin{aligned}\delta\dot{\rho} &= -[\hat{\omega}\times]\,\delta\rho + 1/2\,\delta\omega \\ \delta\dot{q}_4 &= 0,\end{aligned}$$

*In particular, [22] has shown that the 4×4 matrix $\Omega(\bar{\omega}^+_k)$ is simultaneously orthogonal and skew-symmetric.

where the angular velocity estimate is given by $\hat{\omega} = \tilde{\omega} - \hat{\beta}$ and the estimation error of angular velocity is defined as $\delta\omega = -(\delta\beta + \eta_\upsilon)$ so as to include the bias estimation error as $\delta\beta = \beta - \hat{\beta}$ and the measurement noise, η_υ. Therefore, the bias estimation dynamics are governed by

$$\delta\dot{\beta} = \eta_u \tag{17}$$

with the state definitions $\Delta\mathbf{x}^T = [\delta\mathbf{q}^T \ \delta\beta^T]$, the equations can be assembled as

$$\Delta\dot{\mathbf{x}} = \mathbf{F}\,\Delta\mathbf{x} + \mathbf{G}\,\mathbf{w} \tag{18}$$

where

$$\mathbf{F} = \begin{bmatrix} -[\hat{\omega}\times] & 0_{3\times1} & -\frac{1}{2}\mathbf{I}_{3\times3} \\ \mathbf{0}_{1\times3} & 0 & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} \end{bmatrix} \text{ and } \quad \mathbf{G} = \begin{bmatrix} -\frac{1}{2}\mathbf{I}_{3\times3} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{1\times3} & 1 & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times1} & \mathbf{I}_{3\times3} \end{bmatrix} \tag{19}$$

and where $\mathbf{0}_{m\times n}$ denotes the $m \times n$ matrix of zeros. The process noise vector $\mathbf{w}^T = [\eta_\upsilon^T \ \eta_{q4} \ \eta_u^T] \in \Re^7$ contains the zero mean random variable η_{q4} with standard deviation σ_{q4}. This additional process noise term is introduced to represent the modeling error resulting from the linearization process. Ideally, η_{q4} cannot be independent of the other three elements, as it has to lie on the quaternionic geodesic. This insight is particularly useful in tuning the filter. At a given measurement epoch, t_k, multiple measurements can be concatenated as

$$\mathbf{y}_k = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix}_{t_k} = \begin{bmatrix} \mathbf{A}(\delta\mathbf{q})\,\mathbf{A}(\hat{\mathbf{q}}^-)\,\mathbf{r}_1 \\ \mathbf{A}(\delta\mathbf{q})\,\mathbf{A}(\hat{\mathbf{q}}^-)\,\mathbf{r}_2 \\ \vdots \\ \mathbf{A}(\delta\mathbf{q})\,\mathbf{A}(\hat{\mathbf{q}}^-)\,\mathbf{r}_n \end{bmatrix}_{t_k} = \begin{bmatrix} \upsilon_1 \\ \upsilon_2 \\ \vdots \\ \upsilon_n \end{bmatrix}_{t_k} \tag{20}$$

Conforming to the EKF tradition [10], the measurement equations are linearized around the *a priori* estimate of the state. The state itself is not estimated, but the deviation $\Delta\mathbf{x}$ is estimated instead. The quaternion deviation must satisfy a unit norm constraint that will be achieved with the CKF algorithm.

Numerical Results

In this section, the performance of the proposed filter is investigated. A comparison is also performed with three different filter implementations pertinent to the attitude estimation problem. The first implementation is the standard MEKF [3] (or simply denoted as EKF) implemented from Crassidis and Junkins [10]. The second estimator is the EKF with linearized state variable equality constraints, as developed originally by Simon and Chia [7]. The third estimator is the Unscented Attitude Filter developed by Crassidis and Markley [24]. The numerical comparison presented herein is between estimators presenting local solutions to the nonlinear problem of attitude estimation. Results from global solutions such as the one proposed by Psiaki [8] are not included.

Case 1

Suppose a spacecraft is spinning at the constant angular velocity of $\omega^T(t_0) = [1, 0, 1]$ revolution/day. Measurements consist of six directions to stars provided at the rate of 1 Hz, with the rate integrating gyro operating at the same frequency. A star tracker simulator [25] is used to generate the six direction measurements at the required measurement noise intensity (for this case, $\sigma = 10^{-4}$). At every measurement time, the unit vectors $\mathbf{b}_j$ and the reference vectors $\mathbf{r}_j$ are available. The true initial attitude for cases 1, 2, and 3 of the simulations is assumed to be the unit quaternion $\mathbf{q}(t_0) = [0, 0, 0, 1]^T$.

The estimated initial quaternion (also for cases 2 and 3) to initialize the Kalman filter is assumed to be $\hat{\mathbf{q}}(t_0) = [1, 0, 0, 0]^T$ a principal angle of π away from the truth described about the principal axis $\mathbf{e} = \{1, 0, 0\}^T$). The two process noise parameters are $\sigma_u = \sqrt{10}\ 10^{-10}$ rad/s$^{3/2}$ and $\sigma_\upsilon = \sqrt{10}\ 10^{-7}$ rad/s$^{3/2}$, respectively. The initial unknown bias of the rate gyro is $\beta(t_0) = [1 \ 1 \ 1]^T$ deg/h. The initial error standard deviation $\sigma_\rho(t_0)$ is set at 0.1 deg uniformly among all three channels of the vector component of the quaternion and 0.2 deg/h for angular rate components, $\sigma_\beta(t_0)$. The uncertainty along the fourth component of the quaternion $\sigma_{q4}(t_0)$ is set to 0.5176 rad considering a small angle assumption and bearing in mind that the trace of the covariance matrix is bounded near unity due to the norm constraint. The initial bias estimate is $\hat{\beta}^T(t_0) = [20.6, 41.2, 41.2]$deg/h. The initial bias estimate is chosen outside the initial bias estimation error covariance levels to stress the algorithms. Some of the parameters above do not exist for the EKF and the Unscented Kalman Filter (UKF). The filter parameters for the UKF (refer to Crassidis and Markley for more information [24]) are set at $a = 1$ and $\lambda = -3$ for cases 1, 2, and 3. To keep the comparisons consistent, the tuning parameters are set at values where the EKF performs in a well-tuned manner. This choice freezes all the tuning parameters except for the ones corresponding to the fourth component of the quaternion (initial uncertainty and the process noise level), which are parameters of much lower sensitivity when compared to all the others. In other words, once the 'classical' tuning parameters are set to some nominal values, there is not much freedom left in the tuning part of the new filter. This is important since the analyst does not have to spend significant effort in tuning the additional degrees of freedom realized by the constraint handling mechanism developed in this paper. The observations made are valid only for the situations discussed here. The nonlinearity of the problem makes it difficult to extrapolate to other problems, and hence the analyst has to investigate the observations on a case by case basis.

The filter is run for 100 times and the average attitude errors are computed from these data (Monte Carlo simulations). The vector component of the error quaternion is employed for error computations, as, for small angles, the magnitude of the vector component of the error quaternion is known to represent the attitude error sufficiently well. Figure 1 shows the average attitude errors incurred by each of the estimators over a 100 different measurement sets.

From Figure 1, it is clear that the linear constrained filter (denoted by LCKF1 in the legend) has trouble converging. The other filter architectures seem to be converging to approximately 10 deg of error in the 2.5-h simulation.

Case 2

A considerably different case is demonstrated using a different value for the true constant angular velocity $\hat{\omega} = [10, 0, 0]^T$ revolutions/day. The process noise is inflated ($\sigma_u = \sqrt{10}\ 10^{-8}$ rad/s$^{3/2}$ and $\sigma_\upsilon\ \sqrt{10}\ 10^{-5}$ rad/s$^{3/2}$ to handle this change. The comparison of the attitude errors incurred by the different filters (average of 100 runs of a Monte Carlo simulation) is presented in Figure 2. It is clear from Figure 2 that for this case, all the filters perform equally well. Convergence is quite good.

Case 3

Case 3 has been devised to evaluate the algorithm's performance in the presence of noisy measurements. The standard deviation for the additive vector white noise to the measurements is increased to $\sigma_{\mathbf{v}j} = 10^{-2}$. In the presence of increased measurement noise, the convergence time of all the filters increased. This coincidentally conforms to the intuitive learning theory rule that states that in the presence of a noisy environment, one should not learn too fast. All the filters seem to slowly agree with each other toward the end of the 27-h simulation. The simulation parameters for cases 1, 2, and 3 are provided in Table 1. To facilitate the longer simulation run, the sampling rate is decreased and hence the algorithms are stressed further in the absence of rich measurement sampling rates. The results are shown in Figure 3.

TRMM Example of Crassidis and Markley

The performance of the filters is compared using a realistic spacecraft model. The TRMM spacecraft is a representative Earth-pointing spacecraft in a near-circular 90-min (350-km) orbit with an inclination of 35 deg. The spacecraft is assumed to be equipped with three-axis magnetometers and gyroscopic rate sensors whose specifications are provided in [24].

The parameters for the UKF are $a = 1$, $\lambda = 0$. A true orientation, given by the quaternion $\mathbf{q}_0 = [-0.0111, 0.7070, 0.5855, 0.3964]^T$ is considered. Similar to [24], two situations of interest are investigated in the comparisons. An initial quaternion estimate set to be at an orientation obtained by rotating the true attitude quaternion by a principal angle of $\pi/2$ along the principal axis of $\mathbf{e} = [001]^T$ is given as the starting estimate in both cases. The first case is executed with initial bias errors set to zero. Results of a single simulation run are shown in Figure 4. A more dramatic situation arises by setting the initial bias estimate to be $\hat{\beta}(t_0) = 10^{-2}\,[1\,2\,1]^T$ rad/s. For all filters to handle this large uncertainty, the initial bias standard deviation is set to $\sigma_{\beta 0}$ = 20 deg/h. The extra tuning parameters for the CKF and the Linear Constrained Kalman filter (LCKF1) are set at $\sigma_{q4,0}$ = 0.7654 rad, and the corresponding process noise level is set to be $\sigma_{q4} = 0.0314$. The results in this case are shown in Figure 5. It is interesting to note that the LCKF1 performs the best in this situation.

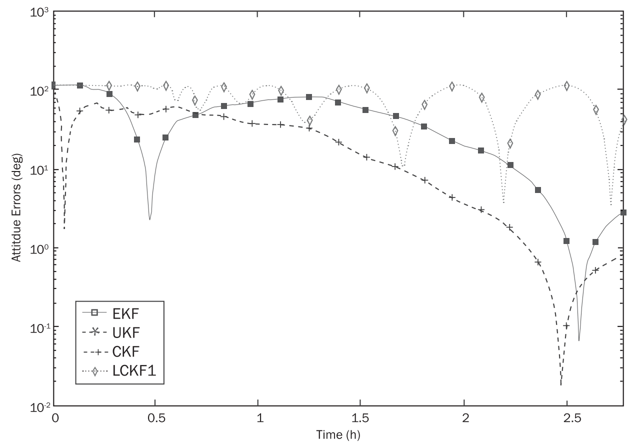


Figure 1. Case 1: Attitude error comparison. EKF denotes MEKF, UKF denotes the unscented Kalman filter, CKF denotes constrained Kalman filter developed in this paper, and LCKF1 denotes the linearized constrained Kalman filter by Simon and Chia.

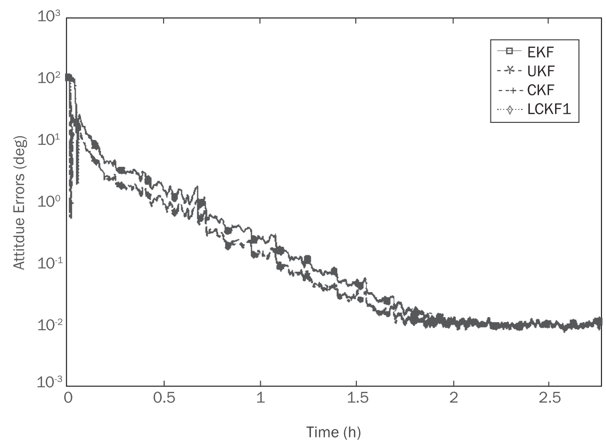


Figure 2. Case 2: Attitude error comparison.

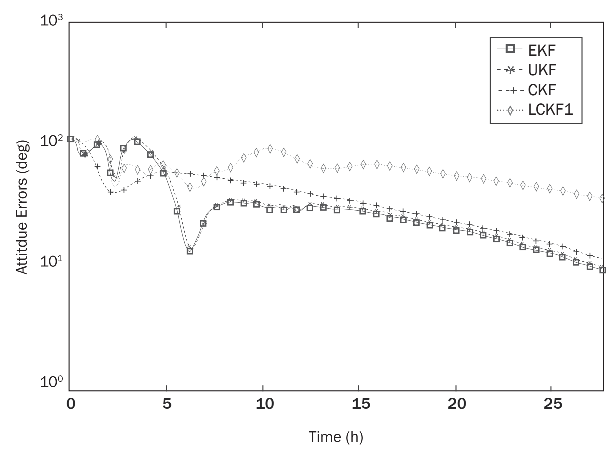


Figure 3. Case 3: Attitude error comparison.

Table 1. Simulation Parameters and Plots Showing the Results.

	Case 1	Case 2	Case 3
Measurement Noise	N(0, 10^{-4})	N(0, 10^{-4})	N(0, 10^{-2})
Process Noise Parameter σ_u (rad/$s^{3/2}$)	$\sqrt{10}\ 10^{-10}$	$\sqrt{10}\ 10^{-8}$	$\sqrt{10}\ 10^{-10}$
Process Noise Parameter σ_v (rad/$s^{3/2}$)	$\sqrt{10}\ 10^{-7}$	$\sqrt{10}\ 10^{-5}$	$\sqrt{10}\ 10^{-7}$
Process Noise Parameter σ_{q4}	1.05×10^{-2}	1.05×10^{-2}	9×10^{-3}
Angular Velocity ω_{true} (rev/day)	$[1, 0, 1]^T$	$[10, 0, 0]^T$	$[1, 0, 0]^T$
Initial Attitude Error (rad)	π	π	π
Initial Bias Estimate $\beta^T(t_0)$	$[1, 2, 2]\ 10^{-4}$	$[1, 2, 2]\ 10^{-4}$	$[1, 2, 2]\ 10^{-4}$
$\sigma_{\rho 0}$ (deg)	1.7×10^{-3}	1.7×10^{-2}	1.7×10^{-3}
$\sigma_{q4,0}$ (rad)	0.5176	0.5176	0.3902
$\sigma_{\beta 0}$ (rad/s)	9.69×10^{-7}	9.69×10^{-6}	9.69×10^{-7}
Simulation Time of Interest (h)	2.7778	2.7778	27.7778
Sampling Frequency (Hz)	1	1	10
Tuning Parameter for UKF [24] *a*	1	1	1
Tuning Parameter for UKF [24] λ	-3	-3	-3
Attitude Error Comparison	Figure 1	Figure 2	Figure 3

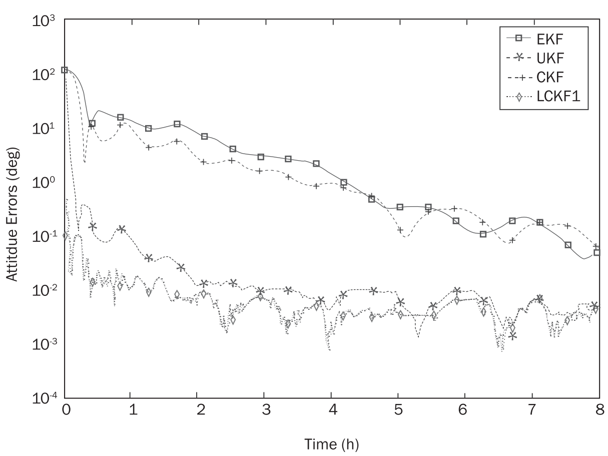


Figure 4. Case 4: Attitude estimation error for case with low initial bias error.

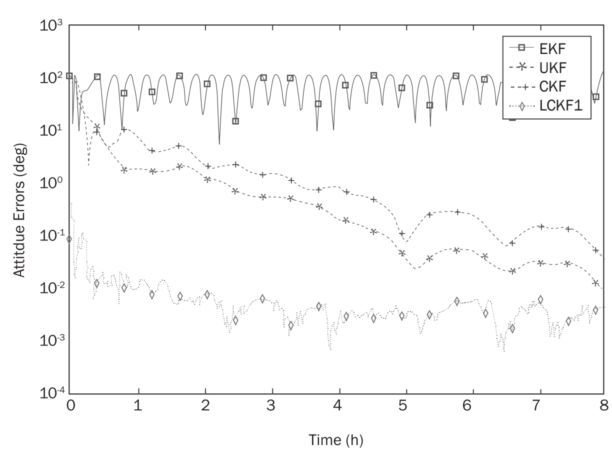


Figure 5. Case 4: Attitude estimation error for the case with large initial bias error.

Conclusions

A novel method to estimate the state vector is presented in this paper for the important case when a subset of the state vector needs to satisfy the constraint of having a given magnitude (e.g., quaternion). The motivation arises from the area of attitude estimation, but the proposed method is general. Examples of the use of the proposed norm-constrained Kalman filter to the attitude estimation problem are provided. In this problem, the state vector consists of a four-component parametrization of the spacecraft orientation (quaternion) and is found to be naturally constrained to have a unit norm. Numerical comparisons with the classical MEKF, the UKF, and the linear constrained Kalman filter have been included for three different typical scenarios.

Appendix

The following calculus identities are used in the derivations:

$d/d\mathbf{X}\,(\mathbf{a}^T\mathbf{X}^T\mathbf{b}) = \mathbf{b}\mathbf{a}^T$

$d/d\mathbf{X}\,(\mathbf{a}^T\mathbf{X}\mathbf{b}) = \mathbf{a}\mathbf{b}^T$

$d/d\mathbf{X}\,(\mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{b}) = \mathbf{X}(\mathbf{a}\mathbf{b}^T + \mathbf{b}\mathbf{a}^T)$

$d/d\mathbf{X}\,(\text{trace}(\mathbf{A}^T\mathbf{X}\mathbf{B}^T)) = d/d\mathbf{X}(\text{trace}(\mathbf{B}\mathbf{X}^T\mathbf{A})) = \mathbf{A}\mathbf{B}$

$d/d\mathbf{X}\,(\text{trace}(\mathbf{X}\mathbf{A}\mathbf{X}^T)) = \mathbf{X}(\mathbf{A} + \mathbf{A}^T)$

Capital bold letters indicate matrices, lowercase bold indicates column vector. The matrix inversion lemma is also used.

$$if \quad det(\mathbf{C} + \mathbf{VAU}) \neq 0 \Rightarrow (\mathbf{A}^{-1} + \mathbf{UC}^{-1}\mathbf{V})^{-1} = \mathbf{A} - \mathbf{AU}(\mathbf{C} + \mathbf{VAU})^{-1}\mathbf{VA}$$

References

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

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

[3] Lefferts, E.J., 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.

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

[5] Richards, P.W., "Constrained Kalman Filtering Using Pseudo-Measurements," *IEE Colloquium on Algorithms for Target Tracking*, May 16, 1995, pp. 75-79.

[6] Lawson, C.L. and R.J. Hanson, "Solving Least Squares Problems," Chap. 20-22, *Automatic Computation*, 1st ed., Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1974.

[7] Simon, D. and T.L. Chia, "Kalman Filtering with State Equality Constraints," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 38, No. 1, January 2002, pp. 128-136.

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

[9] Bar-Itzhack, I.Y. and Y. Oshman, "Attitude Determination from Vector Observations: Quaternion Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 21, No. 1, January 1985, pp. 128-135.

[10] Crassidis, J.L. and J.L. Junkins, "Optimal Estimation of Dynamic Systems," Chap. 7, *Applied Mathematics and Nonlinear Science*, 1st ed., Chapman & Hall/VRV, 2004.

[11] Bar-Itzhack, I.Y., "Optimum Normalization of a Computed Quaternion of Rotation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 7, No. 2, March 1971, pp. 401-402.

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

[13] Bar-Itzhack, I.Y. and J.K. Thienel, "On the Singularity in the Estimation of the Quaternion-of-Rotation," Guidance, Navigation, and Control Conference and Exhibit, AIAA, Monterrey, CA, Paper No. AIAA-2002-4831, August 2002.

[14] Zanetti, R. and R.H. Bishop, "Quaternion Estimation and Norm-Constrained Kalman Filtering," AIAA/AAS, Keystone, Colorado, August 21-24, 2006.

[15] Julier, S.J. and J.J. La Viola Jr., "On Kalman Filtering with Nonlinear State Equality Constraints," *IEEE Transactions on Signal Processing*, Vol. 55, No. 6, June 2007, pp. 103-120.

[16] Choukroun, D., I.Y. Bar-Itzhack, and Y. Oshman, "A Novel Quaternion Kalman Filter," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 42, No. 1, 2006, pp. 174-190.

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

[18] Schaub, H. and J.L. Junkins, "Analytical Mechanics of Space Systems," Chap. 3, *AIAA Education Series*, AIAA, 1st ed., Reston, VA, 2004.

[19] Wertz, J.R., "Spacecraft Attitude Determination and Control," Chap. *Parameterizations of the Attitude* by F.L. Markley, D. Reidel Publishing Co., Dordrecht, Holland, 1978.

[20] Mortari, D., "On the Rigid Rotation Concept in n-Dimensional Spaces," *Journal of the Astronautical Sciences*, Vol. 49, No. 3, July-September 2001, pp. 401-420.

[21] Mortari, D., Scuro, S.R., and Bruccoleri, C., "Attitude and Orbit Error in any Dimensional Spaces," *Journal of the Astronautical Sciences*, Vol. 54, No. 3-4, July-December 2006, pp. 467-484.

[22] Majji, M. and D. Mortari, "Quaternion Constrained Kalman Filter," AAS/AIAA Space Flight Mechanics Meeting, Galveston, Texas, January 27-31, 2008.

[23] Farrenkopf, R., "Analytic Steady-State Accuracy Solutions for Two Common Spacecraft Attitude Estimators," *Journal of Guidance, Control, and Dynamics*, Vol. 1, No. 4, 1978, pp. 282-284.

[24] Crassidis, J.L. and F.L. Markley, "Unscented Filtering for Spacecraft Attitude Estimation," *Journal of Guidance Control and Dynamics*, Vol. 26, No. 4, July-August 2003, pp. 536-541.

[25] Samaan, M.A., D. Mortari, and J.L.Junkins, "Recursive Mode Star Identification Algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 41, No. 4, 2005, pp. 1246-1254.

Renato Zanetti joined Draper Laboratory in October 2007 and has been involved in various space vehicles GN&C projects. His technical interests include stochastic and nonlinear estimation, navigation of aerospace vehicles, autonomous spacecraft rendezvous, and attitude dynamics, determination, and control. He obtained a Ph.D. in Aerospace Engineering at the University of Texas at Austin.

Manoranjan Majji is a Post-Doctoral Research Assistant in the Aerospace Engineering Department at Texas A&M University. His research interests are in the areas of estimation of dynamical systems, system identification, photogrammetry and computational vision. He received a Ph.D. in Aerospace Engineering from Texas A&M University.

Robert H. Bishop is Dean of Engineering at Marquette University. Previously, he was Professor of Aerospace Engineering and Engineering Mechanics at the University of Texas at Austin. He held the Joe J. King Professorship in Engineering, and had served on the Cockrell School of Engineering faculty since 1990. He specializes in GN&C of aerospace vehicles. He has 10 years of industrial experience and has published numerous journal articles and three books. He was the recipient of the 1997 Lockheed Martin Tactical Aircraft Systems Excellence in Engineering Teaching Award, the 1999 John Leland Atwood Award, and was elected to The UT Academy of Distinguished Teachers in 2002. Dr. Bishop was elected Fellow of the American Institute of Aeronautics and Astronautics in 2007. He received a Ph.D. from Rice University.

Daniele Mortari is an Associate Professor of Aerospace Engineering at Texas A&M University. He has made significant contributions in both theory and practice to his field in two key areas of specialization: spacecraft attitude estimation and spacecraft formation design. He taught at the Aerospace Engineering School of Rome's University and was visiting a professor of Electronic Engineering at Perugia University, Italy. He is an associate editor for *AAS Journal of the Astronautical Sciences, International Journal of Navigation and Observations, IEEE Transactions on Aerospace and Electronic Systems*; and an AIAA Associate Fellow, and an IEEE Distinguished Speaker. He has published extensively and holds multiple U.S. patents. Recognition for his work includes two NASA's Group Achievement Awards, the 2003 Spacecraft Technology Center Award, and the prestigious 2007 IEEE Judith A. Resnik Award. Dr. Mortari received a doctoral degree in Nuclear Engineering from the La Sapienza in Rome, Italy.