Joseph Formulation of Unscented and Quadrature Filters with Application to Consider States

Renato Zanetti[1]
The Charles Stark Draper Laboratory, Houston, Texas 77058

Kyle J. DeMars[2]
Air Force Research Laboratory, Space Vehicles Directorate, Kirtland AFB, NM 87117

I. Introduction

The Joseph formula [1] is a general covariance update equation valid not only for the Kalman gain, but for any linear unbiased estimator under standard Kalman filtering assumptions. The Joseph formula is given by $\mathbf{P}^+ = (\mathbf{I} - \mathbf{KH})\mathbf{P}^-(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}}$, where $\mathbf{I}$ is the identity matrix, $\mathbf{K}$ is the gain, $\mathbf{H}$ is the measurement mapping matrix, $\mathbf{R}$ is the measurement noise covariance matrix, and $\mathbf{P}^-$, $\mathbf{P}^+$ are the pre and post measurement update estimation error covariance matrices, respectively. The optimal linear unbiased estimator (equivalently the optimal linear minimum mean square error estimator) or Kalman filter often utilizes simplified covariance update equations such as $\mathbf{P}^+ = (\mathbf{I}-\mathbf{KH})\mathbf{P}^-$ and $\mathbf{P}^+ = \mathbf{P}^- - \mathbf{K}(\mathbf{HP}^-\mathbf{H}^{\mathrm{T}} + \mathbf{R})\mathbf{K}^{\mathrm{T}}$. While these alternative formulations require fewer computations than the Joseph formula, they are only valid when $\mathbf{K}$ is chosen as the optimal Kalman gain. In engineering applications, situations arise where the optimal Kalman gain is not utilized and the Joseph formula must be employed to update the estimation error covariance. Two examples of such a scenario are underweighting measurements [2] and considering states [3]. Even when the optimal gain is used, the Joseph formulation is still preferable because it possesses greater numerical accuracy than the simplified equation [4].

In this note, an equivalent to the Joseph formula is derived for linear estimators but without the assumption of linear measurements. The formula is applied to the quadrature filter [5] and the unscented filter [6] in

---

[1] Senior Member of the Technical Staff, Vehicle Dynamics and Control, 17629 El Camino Real, Suite 470, rzanetti@draper.com, AIAA Senior Member.
[2] National Research Council Postdoctoral Research Fellow, AIAA Member

the presence of consider parameters.

Schmidt's approach for consider states (Schmidt-Kalman filter) is based on minimum variance estimation [7]. Jazwinski [8] details the derivation of the consider Kalman filter in the presence of linear measurements. For nonlinear measurements, the standard extended Kalman filter approach is used, i.e. linearization around the conditional expectation is performed.

Woodbury and Junkins [3] performed a careful analysis of both the Schmidt-Kalman filter and the consider analysis approach as derived by Tapley *et al.* [9]. The analysis by Woodbury and Junkins shows the differences and the benefits of each of the two approaches. The consider filter has received considerable attention in recent years. Woodbury *et al.* provide new insight into considering parameters in the measurement model [10]. Equivalent formulations to the consider filter were also studied [11, 12] and applied to Mars entry navigation [13] and orbit determination [14]. Lisano [15] introduced an unscented formulation of the covariance analysis approach by Tapley *et al.* As described by Woodbury and Junkins that approach is different from that of the Schmidt-Kalman filter.

Instead of deriving the consider filter for linear measurements and then extend the results to nonlinear measurements, this work derives the general linear consider optimal filter in the presence of nonlinear measurements. The optimal estimator reduces to the consider filter in the case of linear measurements and it can be approximated by linearization around the conditional mean to obtain the well known consider filter results. However, this work does not approximate the general consider filter equations via linearization around the mean, but through the use of a set of deterministic points. Depending on the scheme chosen for the points selection, the consider quadrature filter and the consider unscented filter are obtained.

II. Generalized Joseph Formula and Linear Minimum Mean Square Consider Filter

Given an n_x-dimensional random vector $\mathbf{x}$, the mean is denoted by $\mathbf{m_x} \triangleq \mathrm{E}\{\mathbf{x}\}$, and the covariance is denoted by $\mathbf{P_{xx}} \triangleq \mathrm{E}\{(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^{\mathrm{T}}\}$. Additionally, given an n_y-dimensional random vector $\mathbf{y}$, the covariance between $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{P_{xy}} \triangleq \mathrm{E}\{(\mathbf{x}-\mathbf{m_x})\,(\mathbf{y}-\mathbf{m_y})^{\mathrm{T}}\}$.

Let $\mathbf{x}$ be the random vector to be estimated and $\mathbf{y}$ be a random vector whose samples are available; $\mathbf{y}$ is potentially a nonlinear function of $\mathbf{x}$, as well as other non-estimated random states $\mathbf{c}$, and zero-mean white

noise $\mathbf{v}$. Thus, in general, $\mathbf{y}$ may be of the form

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{c}, \mathbf{v})\,.$$

The linear estimators of $\mathbf{x}$ from $\mathbf{y}$ is the family of functions given by $\hat{\mathbf{x}} = \boldsymbol{\ell}(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}$. The goal is to find optimal values for $\mathbf{A}$ and $\mathbf{b}$ in a minimum mean square error (MMSE) sense. The optimal coefficients are denoted with an asterisk. The orthogonality principle [16] is valid when the family of estimation functions is closed under addition and multiplication by a scalar. Under this hypothesis the orthogonality principle establishes that the optimal estimation error, $\mathbf{e} = \mathbf{x} - (\mathbf{A}^*\mathbf{y} + \mathbf{b}^*)$, is perpendicular to every possible estimator, i.e.

$$\mathrm{E}\left\{\left[\mathbf{x} - \mathbf{A}^*\mathbf{y} - \mathbf{b}^*\right]^{\mathrm{T}}\,\left[\mathbf{A}\mathbf{y} + \mathbf{b}\right]\right\} = 0 \;\; \forall\, \mathbf{A}, \mathbf{b} \tag{1}$$

$$\mathbf{b}^{\mathrm{T}}\mathrm{E}\left\{\mathbf{x} - \mathbf{A}^*\mathbf{y} - \mathbf{b}^*\right\} + \mathrm{trace}\left(\mathbf{A}\mathrm{E}\left\{\mathbf{y}\left[\mathbf{x} - \mathbf{A}^*\mathbf{y} - \mathbf{b}^*\right]^{\mathrm{T}}\right\}\right) = 0 \;\; \forall\, \mathbf{A}, \mathbf{b}. \tag{2}$$

Noting that the orthogonality condition must be satisfied for all $\mathbf{A}$ and $\mathbf{b}$ it follows that the coefficients of $\mathbf{b}$ and $\mathbf{A}$ in Eq. (2) must be zero

$$\mathrm{E}\left\{\mathbf{x} - \mathbf{A}^*\mathbf{y} - \mathbf{b}^*\right\} = \mathbf{0} \tag{3}$$

$$\mathrm{E}\left\{\mathbf{y}\left[\mathbf{x} - \mathbf{A}^*\mathbf{y} - \mathbf{b}^*\right]^{\mathrm{T}}\right\} = \mathbf{O} \tag{4}$$

The first condition implies $\mathbf{b}^* = \mathrm{E}\{\mathbf{x}\} - \mathbf{A}^*\mathrm{E}\{\mathbf{y}\} = \mathbf{m}_{\mathbf{x}} - \mathbf{A}^*\mathbf{m}_{\mathbf{y}}$. The linear MMSE (LMMSE) estimator therefore has the form $\hat{\mathbf{x}} = \mathbf{m}_{\mathbf{x}} + \mathbf{A}^*(\mathbf{y} - \mathbf{m}_{\mathbf{y}})$, from which it is established that the estimate is unbiased (i.e. the estimation error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ is zero mean). Combining Eq. (3) and Eq. (4) we obtain that for any vector $\mathbf{m}$ of appropriate dimensions

$$\mathrm{E}\left\{(\mathbf{y} - \mathbf{m})\left[\mathbf{x} - \mathbf{A}^*\mathbf{y} - \mathbf{b}^*\right]^{\mathrm{T}}\right\} = \mathbf{O} \;\; \forall\, \mathbf{m} \tag{5}$$

The optimal gain $\mathbf{A}^*$ can be derived by substituting the optimal $\mathbf{b}^* = \mathbf{m}_{\mathbf{x}} - \mathbf{A}^*\mathbf{m}_{\mathbf{y}}$ into Eq. (5) to obtain

$$\mathrm{E}\left\{(\mathbf{y} - \mathbf{m}_{\mathbf{y}})\left[(\mathbf{x} - \mathbf{m}_{\mathbf{x}}) - \mathbf{A}^*(\mathbf{y} - \mathbf{m}_{\mathbf{y}})\right]^{\mathrm{T}}\right\} = \mathbf{O}\,,$$

the optimal matrix is therefore given by

$$\mathbf{A}^* = \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}\,, \tag{6}$$

where $\mathbf{P}_{\mathbf{yy}}^{-1}$ is the matrix inverse of $\mathbf{P}_{\mathbf{yy}}$. The LMMSE estimator is therefore given by

$$\hat{\mathbf{x}} = \mathbf{m}_{\mathbf{x}} + \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}\left(\mathbf{y} - \mathbf{m}_{\mathbf{y}}\right) . \tag{7}$$

When introducing consider states $\mathbf{c}$, it is necessary to know their covariance and the correlation between them and $\mathbf{x}$ in order to calculate $\mathbf{P}_{\mathbf{xy}}$ and $\mathbf{P}_{\mathbf{yy}}$. When measurements are linear and in the absence of consider states

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

$$\mathbf{P}_{\mathbf{xy}} = \mathbf{P}_{\mathbf{xx}}\mathbf{H}^{\mathrm{T}}$$

$$\mathbf{P}_{\mathbf{yy}} = \mathbf{H}\mathbf{P}_{\mathbf{xx}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\,,$$

where $\mathbf{R}$ is the covariance of the zero-mean measurement noise $\mathbf{v}$. When substituting the above equations in Eq. (7) the familiar Kalman filter emerges.

The family of all linear unbiased estimators is given by $\hat{\mathbf{x}} = \mathbf{m}_{\mathbf{x}} + \mathbf{A}(\mathbf{y} - \mathbf{m}_{\mathbf{y}})$ and their estimation error has covariance matrix $\mathbf{P}_{\mathbf{ee}}$ given by

$$\mathbf{P}_{\mathbf{ee}} = \mathbf{P}_{\mathbf{xx}} - \mathbf{P}_{\mathbf{xy}}\mathbf{A}^{\mathrm{T}} - \mathbf{A}\mathbf{P}_{\mathbf{xy}}^{\mathrm{T}} + \mathbf{A}\mathbf{P}_{\mathbf{yy}}\mathbf{A}^{\mathrm{T}} . \tag{8}$$

Eq. (8) is the equivalent to the Joseph formula in the case of nonlinear measurements; the equation is valid for any value of $\mathbf{A}$, not just the optimal value. When measurements are linear and in the absence of consider states Eq. (8) reduces to the familiar Joseph formula

$$\mathbf{P}_{\mathbf{ee}} = (\mathbf{I} - \mathbf{A}\mathbf{H})\mathbf{P}_{\mathbf{xx}}(\mathbf{I} - \mathbf{A}\mathbf{H})^{\mathrm{T}} + \mathbf{A}\mathbf{R}\mathbf{A}^{\mathrm{T}}.$$

In the presence of nonlinear measurements and consider states, we define an augmented state vector $\mathbf{z}^{\mathrm{T}} = [\mathbf{x}^{\mathrm{T}}\ \mathbf{c}^{\mathrm{T}}]$, and the linear consider estimator is given by

$$\hat{\mathbf{z}} = \mathbf{b} + \mathbf{K}_{\mathrm{con}}\mathbf{y}\,,$$

where the rows of $\mathbf{K}_{\mathrm{con}}$ corresponding to $\mathbf{c}$ are zero. The family of all linear consider estimators is closed under addition and multiplication by a scalar, therefore the orthogonality principle holds, and the same steps previously used in determining optimal values for $\mathbf{b}$ and $\mathbf{K}_{\mathrm{con}}$ can be repeated to obtain the optimal consider state update

$$\hat{\mathbf{z}} = \mathbf{m}_{\mathbf{z}} + \mathbf{K}_{\mathrm{con}}(\mathbf{y} - \mathbf{m}_{\mathbf{y}}) = \begin{bmatrix} \mathbf{m}_{\mathbf{x}} \\ \mathbf{m}_{\mathbf{c}} \end{bmatrix} + \begin{bmatrix} \mathbf{A}^{*} \\ \mathbf{O} \end{bmatrix} (\mathbf{y} - \mathbf{m}_{\mathbf{y}}) = \begin{bmatrix} \mathbf{m}_{\mathbf{x}} + \mathbf{A}^{*}(\mathbf{y} - \mathbf{m}_{\mathbf{y}}) \\ \mathbf{m}_{\mathbf{c}} \end{bmatrix} , \tag{9}$$

where $\mathbf{A}^*$ is defined in Eq. (6). The update of the estimation error covariance is given by the generalized Joseph formula

$$\mathbf{P}_{\mathrm{aug}} = \mathbf{P}_{\mathbf{zz}} - \mathbf{P}_{\mathbf{zy}}\mathbf{K}_{\mathrm{con}}^{\mathrm{T}} - \mathbf{K}_{\mathrm{con}}\mathbf{P}_{\mathbf{zy}}^{\mathrm{T}} + \mathbf{P}_{\mathbf{zy}}\mathbf{P}_{\mathbf{yy}}\mathbf{P}_{\mathbf{zy}}^{\mathrm{T}}\,. \tag{10}$$

For linear measurements and consider states Eqs. (9) and (10) reduce to the consider filter.

III. New Consider Filter Algorithms

In order to implement the consider filter that is described by Eqs. (9) and (10), the values of $\mathbf{m}_{\mathbf{y}}$, $\mathbf{P}_{\mathbf{yy}}$, $\mathbf{P}_{\mathbf{xy}}$, and $\mathbf{P}_{\mathbf{zy}}$ need to be determined. First, define a composite input, $\mathbf{u}$, to the measurement function such that $\mathbf{u}^{\mathrm{T}} = [\mathbf{x}^{\mathrm{T}}\,\mathbf{c}^{\mathrm{T}}\,\mathbf{v}^{\mathrm{T}}]$. Given $\mathbf{x} \in \mathbb{R}^{n_x}$, $\mathbf{c} \in \mathbb{R}^{n_c}$, and $\mathbf{v} \in \mathbb{R}^{n_v}$, it follows that $\mathbf{u} \in \mathbb{R}^n$ where $n = n_x + n_c + n_v$, and that the measurement function may be expressed as

$$\mathbf{y} = \mathbf{h}(\mathbf{u})\,.$$

Recalling that $\mathbf{y} \in \mathbb{R}^{n_y}$ and given a value of $\mathbf{P}_{\mathbf{uy}}$, it follows that $\mathbf{P}_{\mathbf{zy}}$ is the upper $(n_x + n_c) \times n_y$ block of $\mathbf{P}_{\mathbf{uy}}$. Furthermore, $\mathbf{P}_{\mathbf{xy}}$ is the upper $n_x \times n_y$ block of $\mathbf{P}_{\mathbf{uy}}$. Therefore, given the values of $\mathbf{m}_{\mathbf{y}}$, $\mathbf{P}_{\mathbf{yy}}$, and $\mathbf{P}_{\mathbf{uy}}$, the necessary components required in Eqs. (9) and (10) are available.

The *a priori* mean and covariance of the composite input, $\mathbf{m}_{\mathbf{u}}$ and $\mathbf{P}_{\mathbf{uu}}$, are known, and are given by

$$\mathbf{m}_{\mathbf{u}} = \begin{bmatrix} \mathbf{m}_{\mathbf{x}} \\ \mathbf{m}_{\mathbf{c}} \\ \mathbf{m}_{\mathbf{v}} \end{bmatrix} \quad \text{and} \quad \mathbf{P}_{\mathbf{uu}} = \begin{bmatrix} \mathbf{P}_{\mathbf{xx}} & \mathbf{P}_{\mathbf{xc}} & \mathbf{P}_{\mathbf{xv}} \\ \mathbf{P}_{\mathbf{cx}} & \mathbf{P}_{\mathbf{cc}} & \mathbf{P}_{\mathbf{cv}} \\ \mathbf{P}_{\mathbf{vx}} & \mathbf{P}_{\mathbf{vc}} & \mathbf{P}_{\mathbf{vv}} \end{bmatrix},$$

where $\mathbf{P}_{\mathbf{cx}} = \mathbf{P}_{\mathbf{xc}}^{\mathrm{T}}$, $\mathbf{P}_{\mathbf{vx}} = \mathbf{P}_{\mathbf{xv}}^{\mathrm{T}}$, and $\mathbf{P}_{\mathbf{vc}} = \mathbf{P}_{\mathbf{cv}}^{\mathrm{T}}$. For zero-mean noise with covariance $\mathbf{R}$, $\mathbf{m}_{\mathbf{v}} = \mathbf{0}$ and $\mathbf{P}_{\mathbf{vv}} = \mathbf{R}$. Additionally, when the noise is not correlated with the state or consider states, $\mathbf{P}_{\mathbf{xv}} = \mathbf{P}_{\mathbf{vx}}^{\mathrm{T}} = \mathbf{O}$ and $\mathbf{P}_{\mathbf{cv}} = \mathbf{P}_{\mathbf{vc}}^{\mathrm{T}} = \mathbf{O}$.

The *a priori* probability density function of $\mathbf{u}$ is denoted as $p(\mathbf{u})$; from it, the mean, covariance, and cross-covariance are obtained as

$$\begin{aligned} \mathbf{m}_{\mathbf{y}} &= \int_{\mathbb{R}^n} \mathbf{h}(\mathbf{u}) p(\mathbf{u}) d\mathbf{u} \\ \mathbf{P}_{\mathbf{yy}} &= \int_{\mathbb{R}^n} (\mathbf{h}(\mathbf{u}) - \mathbf{m}_{\mathbf{y}})(\mathbf{h}(\mathbf{u}) - \mathbf{m}_{\mathbf{y}})^{\mathrm{T}} p(\mathbf{u}) d\mathbf{u} \\ \mathbf{P}_{\mathbf{uy}} &= \int_{\mathbb{R}^n} (\mathbf{u} - \mathbf{m}_{\mathbf{u}})(\mathbf{h}(\mathbf{u}) - \mathbf{m}_{\mathbf{y}})^{\mathrm{T}} p(\mathbf{u}) d\mathbf{u}\,. \end{aligned} \tag{11}$$

The covariance terms admit a simplification as $\mathbf{P_{yy}} = \tilde{\mathbf{P}}_{\mathbf{yy}} - \mathbf{m_y}\mathbf{m_y}^{\mathrm{T}}$ and $\mathbf{P_{uy}} = \tilde{\mathbf{P}}_{\mathbf{uy}} - \mathbf{m_u}\mathbf{m_y}^{\mathrm{T}}$, where

$$\tilde{\mathbf{P}}_{\mathbf{yy}} = \int_{\mathbb{R}^n} \mathbf{h}(\mathbf{u})\mathbf{h}^{\mathrm{T}}(\mathbf{u})p(\mathbf{u})d\mathbf{u} \tag{12}$$

$$\tilde{\mathbf{P}}_{\mathbf{uy}} = \int_{\mathbb{R}^n} \mathbf{u}\mathbf{h}^{\mathrm{T}}(\mathbf{u})p(\mathbf{u})d\mathbf{u}\,. \tag{13}$$

Therefore, the three integral terms of Eqs. (11)–(13) need to be evaluated in order to evaluate the consider filter that is described by Eqs. (9) and (10), where each of the three terms has the form

$$I = \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{u})p(\mathbf{u})d\mathbf{u}\,; \tag{14}$$

the quadrature and unscented filters approximate these integrals by the summation of a finite number of deterministic points.

A. The Consider Quadrature Kalman Filter

The quadrature Kalman filter assumes that the *a priori* density is Gaussian with mean $\mathbf{m_u}$ and covariance $\mathbf{P_{uu}}$, i.e.

$$p(\mathbf{u}) = |2\pi\mathbf{P_{uu}}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{u}-\mathbf{m_u})^{\mathrm{T}}\mathbf{P_{uu}}^{-1}(\mathbf{u}-\mathbf{m_u})\right\}\,.$$

The method is based on the Gauss-Hermite quadrature rule, which is given by

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} f(u)e^{-u^2}du = \sum_{i=1}^{m} w_i f(q_i)\,,$$

where q_i and w_i are the quadrature points and weights, respectively, and the equality holds for all polynomials of degree up to $2m-1$, where m is the chosen order of the quadrature rule. The quadrature points and weights can be determined via an eigenvalue problem as follows. Let $\mathbf{J}$ be a symmetric, tridiagonal matrix with zeros on the main diagonal. The elements of the first upper and lower diagonals are given by $J_{i,i+1} = J_{i+1,i} = \sqrt{i/2}$ for $1 \leq i \leq m-1$. Then, the quadrature points are the eigenvalues of $\mathbf{J}$ and the quadrature weights are given by $w_i = |(v_i)_1|^2$, where $(v_i)_1$ is the first element of the i^{th} normalized eigenvector of $\mathbf{J}$ [5, 17].

Consider a scalar random variable, u, which is distributed according to a standard normal distribution (i.e. a Gaussian distribution with zero mean and unit variance). It readily follows by a change of variables that the Gauss-Hermite quadrature rule may be employed as

$$\int_{-\infty}^{\infty} f(u)\mathcal{N}(u;0,1)du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(u)e^{-u^2/2}du = \sum_{i=1}^{m} w_i f(\kappa_i)\,,$$

where $\kappa_i = \sqrt{2}\, q_i$. In the case of an n-dimensional vector-valued random variable, $\mathbf{u}'$, with zero mean and identity variance, the univariate Gauss-Hermite quadrature rule is extended to a multivariate quadrature rule by successive application to the mutually uncorrelated elements of $\mathbf{u}'$, yielding [5]

$$\int_{\mathbb{R}^n} \mathbf{f}(\mathbf{u}')\mathcal{N}(\mathbf{u}';\mathbf{0},\mathbf{I})d\mathbf{u}' = \sum_{i_n=1}^{m} w_{i_n} \cdots \sum_{i_1=1}^{m} w_{i_1} \mathbf{f}(\kappa_{i_1},\ldots,\kappa_{i_n}) = \sum_{i=1}^{m^n} \lambda_i \mathbf{f}(\boldsymbol{\kappa}_i)\,,$$

where $\boldsymbol{\kappa}_i = [\kappa_{i_1} \cdots \kappa_{i_n}]^{\mathrm{T}}$ and $\lambda_i = \prod_{j=1}^{n} w_{i_j}$. Thus, an m-point univariate quadrature rule generates an m^n-point quadrature rule for n-dimensional integral evaluations. While the previous equation represents an n-dimensional quadrature, it is not of the form expressed in Eq. (14). Since an arbitrary multivariate Gaussian distribution is a linear transformation from a zero-mean, unit-variance Gaussian distribution, the final step is to perform a linear change of variables, which yields

$$\int_{\mathbb{R}^n} \mathbf{f}(\mathbf{u})\mathcal{N}(\mathbf{u};\mathbf{m_u},\mathbf{P_{uu}})d\mathbf{u} = \sum_{i=1}^{m^n} \lambda_i \mathbf{f}(\boldsymbol{\mathcal{U}}_i)\,, \tag{15}$$

where $\boldsymbol{\mathcal{U}}_i = \mathbf{m_u} + \mathbf{S_{uu}}\boldsymbol{\kappa}_i$ and $\mathbf{S_{uu}}$ is a square-root factor of $\mathbf{P_{uu}}$, such that $\mathbf{P_{uu}} = \mathbf{S_{uu}}\mathbf{S_{uu}}^{\mathrm{T}}$.

In order to utilize the quadrature approach for the consider filter, first select the quadrature rule via the parameter m. Using the previously described approach, generate the n-dimensional quadrature rule, yielding the m^n quadrature points $\boldsymbol{\kappa}_i$ and associated weights λ_i. Compute the square-root factor $\mathbf{S_{uu}}$ from $\mathbf{P_{uu}}$ (e.g. using a Cholesky factorization) in order to determine $\boldsymbol{\mathcal{U}}_i = \mathbf{m_u} + \mathbf{S_{uu}}\boldsymbol{\kappa}_i$. Then, the integral terms of Eqs. (11)–(13) are computed via Eq. (15) as

$$\begin{aligned}
\mathbf{m_y} &= \sum_{i=1}^{m^n} \lambda_i \mathbf{h}(\boldsymbol{\mathcal{U}}_i) \\
\tilde{\mathbf{P}}_{\mathbf{yy}} &= \sum_{i=1}^{m^n} \lambda_i \mathbf{h}(\boldsymbol{\mathcal{U}}_i)\mathbf{h}^{\mathrm{T}}(\boldsymbol{\mathcal{U}}_i) \\
\tilde{\mathbf{P}}_{\mathbf{uy}} &= \sum_{i=1}^{m^n} \lambda_i \boldsymbol{\mathcal{U}}_i \mathbf{h}^{\mathrm{T}}(\boldsymbol{\mathcal{U}}_i)\,.
\end{aligned}$$

$\mathbf{P_{yy}}$ and $\mathbf{P_{uy}}$ are then given by $\mathbf{P_{yy}} = \tilde{\mathbf{P}}_{\mathbf{yy}} - \mathbf{m_y}\mathbf{m_y}^{\mathrm{T}}$ and $\mathbf{P_{uy}} = \tilde{\mathbf{P}}_{\mathbf{uy}} - \mathbf{m_u}\mathbf{m_y}^{\mathrm{T}}$, from which $\mathbf{P_{zy}}$ and $\mathbf{P_{xy}}$ may be extracted. Finally, use Eqs. (9) and (10) to complete the quadrature consider filter.

B. The Consider Unscented Kalman Filter

Given an n-dimensional random variable $\mathbf{u}$ with mean and covariance, $\mathbf{m_u}$ and $\mathbf{P_{uu}}$, respectively, and a nonlinear transformation

$$\mathbf{y} = \mathbf{h}(\mathbf{u})\,,$$

the unscented Kalman filter, like the quadrature Kalman filter, employs a set of deterministically selected points in order to compute the mean and covariance of $\mathbf{y}$, as well as the cross-covariance between $\mathbf{u}$ and $\mathbf{y}$. Unlike the quadrature Kalman filter, the unscented Kalman filter selects its points based on moment matching. That is, a set of sigma-points, $\boldsymbol{\mathcal{U}}_i$ and associated weights, w_i, are selected so that the moments of $\mathbf{y}$ are well approximated. In general, given a set of K sigma-points, $\boldsymbol{\mathcal{U}}_i$, and the transformed values, $\boldsymbol{\mathcal{Y}}_i = \mathbf{h}(\boldsymbol{\mathcal{U}}_i)$, the mean, covariance, and cross-covariance are computed as

$$\mathbf{m_y} = \sum_{i \in \mathcal{I}} w_i^{(m)} \boldsymbol{\mathcal{Y}}_i \tag{16a}$$

$$\tilde{\mathbf{P}}_{\mathbf{yy}} = \sum_{i \in \mathcal{I}} w_i^{(c)} \boldsymbol{\mathcal{Y}}_i \boldsymbol{\mathcal{Y}}_i^{\mathrm{T}} \tag{16b}$$

$$\tilde{\mathbf{P}}_{\mathbf{uy}} = \sum_{i \in \mathcal{I}} w_i^{(c)} \boldsymbol{\mathcal{U}}_i \boldsymbol{\mathcal{Y}}_i^{\mathrm{T}} \,, \tag{16c}$$

with $\mathbf{P_{yy}} = \tilde{\mathbf{P}}_{\mathbf{yy}} - \mathbf{m_y}\mathbf{m_y}^{\mathrm{T}}$ and $\mathbf{P_{uy}} = \tilde{\mathbf{P}}_{\mathbf{uy}} - \mathbf{m_u}\mathbf{m_y}^{\mathrm{T}}$, and where the cardinality of $\mathcal{I}$ is K, i.e. the number of sigma-points. It should be noted that the unscented Kalman filter can employ different weights for the mean and covariance calculations. Three methods for constructing the input sigma-points and their associated weights are reviewed: the symmetric, extended symmetric, and scaled extended symmetric sigma-point selection schemes.

The symmetric sigma-point selection scheme chooses a set of $K = 2n$ sigma-points that are on the $\sqrt{n}^{\mathrm{th}}$ covariance contour as [18]

$$\begin{aligned} \boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} + \sqrt{n}\, \mathbf{s}_i \qquad i = 1, \ldots, n \\ \boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} - \sqrt{n}\, \mathbf{s}_{i-n} \qquad i = n+1, \ldots, 2n \,, \end{aligned}$$

with associated weights of $w_i^{(m)} = w_i^{(c)} = 1/2n$ for $i = 1, \ldots, 2n$, and $\mathcal{I} = \{1, \ldots, 2n\}$. Here, $\mathbf{s}_i$ represents the i^{th} column of the square-root factor of the covariance matrix, i.e. $\mathbf{s}_i$ is the i^{th} column of $\mathbf{S_{uu}}$, where $\mathbf{S_{uu}}\mathbf{S_{uu}}^{\mathrm{T}} = \mathbf{P_{uu}}$.

The symmetric sigma-point selection scheme guarantees matching of the mean and covariance of the input distribution. Additionally, since the scheme is symmetric by construction, the third moment for symmetric distributions is also matched; however, introduction of a tuning parameter (and another sigma-point) enables the sigma-points to capture up to 4^{th} moments. This is done by extending the symmetric sigma-point set to include an additional sigma-point that is the mean, yielding the extended symmetric sigma-point

selection scheme as [19]

$$\begin{aligned}
\boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} & i &= 0 \\
\boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} + \sqrt{n+\kappa}\,\mathbf{s}_i & i &= 1,\ldots,n \\
\boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} - \sqrt{n+\kappa}\,\mathbf{s}_{i-n} & i &= n+1,\ldots,2n\,,
\end{aligned}$$

with weights given by $w_i^{(m)} = w_i^{(c)} = \kappa/(n+\kappa)$ for $i = 0$, and $w_i^{(m)} = w_i^{(c)} = 1/2(n+\kappa)$ for $i = 1,\ldots,2n$, and with $\mathcal{I} = \{1,\ldots,2n+1\}$. Choosing κ such that $n+\kappa = 3$ ensures that the 4^{th} moment matches [19].

When $\kappa = 3 - n < 0$, the weight for $\boldsymbol{\mathcal{U}}_0$ becomes negative, and the calculated covariance can become non-positive semidefinite [20]. This effect motivated the development of the scaled unscented transform which replaces the extended symmetric sigma-points with the scaled extended symmetric set of sigma-points as

$$\boldsymbol{\mathcal{U}}_i' = \boldsymbol{\mathcal{U}}_0 + \alpha(\boldsymbol{\mathcal{U}}_i - \boldsymbol{\mathcal{U}}_0)$$

for $i = 1,\ldots,2n$, where α is a positive scaling parameter such that $0 \leq \alpha \leq 1$. Additionally, since the weighting of the mean sigma-point directly affects the magnitude of the errors in the fourth and higher order terms for symmetric prior distributions, a third parameter, β is introduced to allow for the minimization of higher order errors in the presence of knowledge of the prior distribution. Thus, the scaled extended symmetric sigma-point selection scheme is given by [20]

$$\begin{aligned}
\boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} & i &= 0 \\
\boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} + \sqrt{n+\lambda}\,\mathbf{s}_i & i &= 1,\ldots,n \\
\boldsymbol{\mathcal{U}}_i &= \mathbf{m_u} - \sqrt{n+\lambda}\,\mathbf{s}_{i-n} & i &= n+1,\ldots,2n\,,
\end{aligned}$$

where $\lambda = \alpha^2(n+\kappa) - n$, and the weights are given by $w_i^{(m)} = \lambda/(n+\lambda)$ for $i = 0$, $w_i^{(c)} = \lambda/(n+\lambda) + (1-\alpha^2+\beta)$ for $i = 0$, and $w_i^{(m)} = w_i^{(c)} = 1/2(n+\lambda)$ for $i = 1,\ldots,2n$. Additionally, $\mathcal{I} = \{1,\ldots,2n+1\}$ for the scaled symmetric sigma-point selection scheme.

In contrast to the extended symmetric sigma-point selection scheme, the scaled extended symmetric sigma-point selection scheme has three tuning parameters: κ, α, and β. Choosing $\kappa \geq 0$ guarantees positive semidefiniteness of the covariance matrix, so a good default value is $\kappa = 0$ [20]. Since α controls the spread of the sigma-points, choosing smaller values of α ensures the avoidance of non-local sampling; choosing $\alpha = 1$, however, produces the same set of sigma-points as the extended symmetric method. Finally, β is a

non-negative parameter that can be used to incoporate prior distribution knowledge; in the case that the prior is Gaussian, the optimal choice is $\beta = 2$ [21].

In order to utilize the unscented approach for the consider filter, first select the sigma-point scheme and any associated tuning parameters. Using the square-root factor $\mathbf{S_{uu}}$ of $\mathbf{P_{uu}}$, determine the sigma-points, $\boldsymbol{\mathcal{U}}_i$, and the associated weights, $w_i^{(m)}$ and $w_i^{(c)}$, according to the chosen scheme. After computing the transformed sigma-points via $\boldsymbol{\mathcal{Y}}_i = \mathbf{h}(\boldsymbol{\mathcal{U}}_i)$ for $i \in \mathcal{I}$, the integral terms of Eqs. (11)–(13) are computed using Eqs. (16). $\mathbf{P_{yy}}$ and $\mathbf{P_{uy}}$ are then given by $\mathbf{P_{yy}} = \tilde{\mathbf{P}}_{\mathbf{yy}} - \mathbf{m_y}\mathbf{m_y}^{\mathrm{T}}$ and $\mathbf{P_{uy}} = \tilde{\mathbf{P}}_{\mathbf{uy}} - \mathbf{m_u}\mathbf{m_y}^{\mathrm{T}}$, from which $\mathbf{P_{zy}}$ and $\mathbf{P_{xy}}$ may be extracted. Finally, use Eqs. (9) and (10) to complete the unscented consider filter.

IV. Conclusions

This note introduces a general covariance update equation which is the extension of the well-known Joseph formula for the nonlinear measurements case. This formula can be used in linear estimators for nonlinear measurements that do not rely on linearization around the current estimate; which is the assumption made by the extended Kalman filter. Two estimation schemes that do not rely on linearization centered the current estimate are the unscented Kalman filter and quadrature filters. The proposed generalized Joseph formula is necessary to update the estimation error covariance whenever a non-optimal gain is chosen in the linear unbiased estimator. Various reasons could dictate the need of a non-optimal gain selection. One reason for the utilization of the generalized Joseph formula and a non-optimal gain is detailed in this note: the inclusion of consider states into the linear estimator. The resulting algorithms are the extension of the well-known consider filter to either the unscented transformation or the Gauss-Hermite quadrature rule.

The classic Joseph formula is known to be more numerically stable than the simplified optimal covariance update equation. The proposed generalized Joseph formula is potentially preferable over the standard covariance update of the unscented and quadrature filters even in the presence of optimal gains for the same reason.

References

[1] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processing with Applications to Guidance*. Providence, RI: AMS Chelsea Publishing, 2nd ed., 2005, pp 55–57.

[2] Renato Zanetti, Kyle J. DeMars, and Robert H. Bishop. Underweighting Nonlinear Measurements. *Journal of Guidance, Control, and Dynamics*, 33(5):1670–1675, September–October 2010.

[3] Woodbury, D. and Junkins, J., "On the Consider Kalman Filter," *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, August 2010.

[4] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Vol. 128 of *Mathematics in Sciences and Engineering*. Academic Press, 1978, pp. 96–100.

[5] Arasaratnam, I., Haykin, S., and Elliot, R. J., "Discrete-Time Nonlinear Filtering Algorithms using Gauss-Hermite Quadrature", *Proceedings of the IEEE*, Vol. 95, No. 5, May 2007, pp. 953–977.

[6] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, March 2000.

[7] Schmidt, S. F., "Application of State-Space Methods to Navigation Problems," *Advances in Control Systems*, Vol. 3, 1966, pp. 293–340.

[8] Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Vol. 64 of *Mathematics in Sciences and Engineering*, Academic Press, New York, New York 10003, 1970, p. 285.

[9] Tapley, B. D., Schutz, B. E., and Born, G. H., *Statistical Orbit Determination*, Elsevier Academic Press, 2004, Chapter 6.

[10] Woodbury, D., Majji, M., and Junkins, J., "Considering Measurement Model Parameter Errors in Static and Dynamic Systems," *Proceedings of the George H. Born Symposium*, May 2010.

[11] Zanetti, R., *Advanced Navigation Algorithms for Precision Landing*, Ph.D. thesis, The University of Texas at Austin, Austin,Texas, December 2007.

[12] Zanetti, R. and Bishop, R. H., "Kalman Filters with Uncompensated Biases," *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 1, January–February 2012, pp. 327–330.

[13] Zanetti, R. and Bishop, R. H., "Entry Navigation Dead-Reckoning Error Analysis: Theoretical Foundations of the Discrete-Time Case," *Proceedings of the AAS/AIAA Astrodynamics Specialist Conference held August 19–23, 2007, Mackinac Island, Michigan*, Vol. 129 of *Advances in the Astronautical Sciences*, 2007, pp. 979–994, AAS 07-311.

[14] Hough, M. E., "Orbit Determination with Improved Covariance Fidelity, Including Sensor Measurement Biases," *Journal of Guidance, Control, and Dynamics*, Vol. 34, No. 3, May–June 2011, pp. 903–911.

[15] Lisano, M. E., "Nonlinear Consider Covariance Analysis Using a Sigma-Point Filter Formulation," 29th Annual AAS Guidance and Control Conference, Breckenridge, Colorado, 4–8 February 2006, AAS 06-035.

[16] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1st ed., 1965, pp. 219–220.

[17] Ito, K. and Xiong, K., "Gaussian Filters for Nonlinear Filtering Problems", *IEEE Transactions on Automatic Con-*

trol, Vol. 45, No. 5, May 2000, pp. 910–927.

[18] Uhlmann, J. K., *Simultaneous Map Building and Localization for Real Time Applications*, Ph.D. thesis, University of Oxford, 1994.

[19] Julier, S. J. and Uhlmann, J. K., "Unscented Filtering and Nonlinear Estimation", *Proceedings of the IEEE*, Vol. 92, March 2004.

[20] van der Merwe, R., *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*, Ph.D. thesis, Oregon Health and Science University, 2004.

[21] Julier, S. J., "The Scaled Unscented Transformation", *Proceedings of the American Control Conference*, May 2002.