Recursive Update Filtering for Nonlinear Estimation

Renato Zanetti

***Abstract*—Nonlinear filters are often very computationally expensive and usually not suitable for real-time applications. Real-time navigation algorithms are typically based on linear estimators, such as the extended Kalman filter (EKF) and, to a much lesser extent, the unscented Kalman filter. This work proposes a novel non-linear estimator whose additional computational cost is comparable to $(N-1)$ EKF updates, where N is the number of recursions, a tuning parameter. The higher N the less the filter relies on the linearization assumption. A second algorithm is proposed with a differential update, which is equivalent to the recursive update as N tends to infinity.**

***Index Terms*—Estimation; Kalman filtering; Filtering; recursive Kalman filter; nonlinear estimation**

I. Introduction

THE well known Kalman filter [1], [2] is an optimal estimation algorithm. The optimality holds in terms of minimum mean square error and maximum likelihood estimation under several conditions. These conditions are met when all noises and the initial estimation error are Gaussian, and when the dynamics and measurements are linear; under these conditions the Kalman filter is globally optimal. For linear measurements/dynamics but without Gaussian distributions, the Kalman filter is not globally optimal, but it still is the linear unbiased minimum variance estimator (i.e. is the optimal out of all linear unbiased estimators) [3]. For the general, non-Gaussian, linear case, the designer faces a choice: investigate the actual distribution of the errors and design a globally optimal filter, or settle for a sub-optimal, distribution independent Kalman filter. The Kalman filter does not require complete knowledge of the error distribution, it only requires knowledge of the first two moments. Gaussian sum filters [4] deal with non-Gaussian distributions. Their additional computational cost and complexity however, make them less attractive than the classical Kalman filter for practical real-time applications. The real error distribution is usually not perfectly known, hence the trade is between a suboptimal Kalman filter versus being able to represent the distributions of all errors accurately.

A widely used algorithm in real-time nonlinear estimation is the extended Kalman filter [5] (EKF). The EKF is a nonlinear approximation of the Kalman filter which assumes small estimation errors and approximates them to first order to calculate their covariance matrix. Like the Kalman filter, the EKF is also a linear estimator but relies on the additional assumption that the first order approximation is valid. Algorithms exists that relax both the aforementioned assumptions. Filters with a polynomial update of arbitrary order have been known since the sixties [6], knowledge of moments of the conditional estimation error distribution higher than the second are needed for these updates. Gaussian sum and particle filters have been used for nonlinear problems [7], [8], but they also require the knowledge of the distributions. Much like in the linear measurement case, renouncing to using a linear estimator requires to know the entire error distribution.

Techniques exist to overcome some of the limitations of the EKF linearization assumption. The Gaussian second order filter (GSOF) [9] takes into account second-order terms assuming the error distribution is Gaussian. The iterated extended Kalman filter [5] recursively improves the center of the Taylor series expansion for a better linearization. The unscented Kalman filter [10] (UKF) is able to retain higher-order terms of the Taylor series expansion. Underweighting [11] is an *ad hoc* technique to compensate for the second order effects without actually computing them.

This paper introduces algorithms to incorporate nonlinear measurements into a Kalman filter. The new nonlinear filter overcomes some of the limitations of the EKF without requiring complete knowledge of the error distribution. The main advantage of the proposed scheme is that the update equation is not necessarily linear and outperforms the EKF for nonlinear measurement functions. Possible divergence due to Taylor series truncation is avoided in the proposed filter and the computational cost is not excessive because the filter is not affected by the curse of dimensionality. The new algorithm does not search among all possible updates, but follows the gradient of the nonlinear measurement function. In a similar manner in which the Kalman filter only searches for the minimum variance estimate out of the linear estimators, the new algorithm minimizes the covariance while following the curvature of the nonlinear measurement. The proposed algorithm is not globally optimal and, like the EKF, can be outperformed by other algorithms in certain nonlinear scenarios. The proposed algorithm reduces to the extended Kalman filter for linear measurements.

The EKF update can suffer from two issues that can make its estimated statistics inconsistent with the true distribution of the error. One possible issue is the first-order truncation of the nonlinear measurement function; the proposed algorithm solves this problem. Another possible limitation of the EKF is not addressed by the proposed update scheme. By evaluating the measurement Jacobian at the estimated value, the EKF makes it a random variable. The Kalman gain is a function of the measurement Jacobian and it is also random. In calculating expectations the Jacobian and the Kalman gain are taken outside the expected value, as if they were constants. The EKF covariance update is valid when the Kalman gain is constant across all possible estimates, which is true for the linear case but not true for the nonlinear case. When the Kalman gain varies significantly, the *a posteriori* EKF covariance can be

Renato Zanetti is with the vehicle dynamics and control group of the Charles Stark Draper Laboratory, Houston, Texas.

inadequate to represent the actual estimation error covariance. The UKF solves this issue by choosing various sigma points spread around the distribution and constructing the gain from them.

The paper is organized as follows. First a quick review of minimum mean square error estimation is presented to motivate the work (Section II). The new algorithms are then introduced in Section III, followed by a comparison to existing methods for nonlinear estimation (Section IV). A numerical example is presented in Section V and some conclusions are drawn in Section VI.

II. Linear Minimum Mean Square Error Estimation

Given an n-dimentional random vector $\mathbf{X}$ with components $X_1, \dots X_n$ and probability density function (pdf) $f_{\mathbf{X}}(\mathbf{x})$, the notation $\mathrm{E}_{\mathbf{X}}\{\mathbf{g}(\mathbf{X})\}$ means

$$\mathrm{E}_{\mathbf{X}}\{\mathbf{g}(\mathbf{X})\} \triangleq \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, dx_1 \dots dx_n. \quad (1)$$

The mean is given by $\boldsymbol{\mu}_{\mathbf{X}} \triangleq \mathrm{E}_{\mathbf{X}}\{\mathbf{X}\}$ and the cross-covariance is $\boldsymbol{\Sigma}_{\mathbf{XY}} \triangleq \mathrm{E}_{\mathbf{XY}}\left\{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})^{\mathrm{T}}\right\}$. For scalars the notation σ^2_{XY} is used instead of Σ_{XY}.

Given two random vectors $\mathbf{X}$ and $\mathbf{Y}$, the linear estimators of $\mathbf{X}$ from $\mathbf{Y}$ are given by $\hat{\mathbf{X}} = \mathbf{A}\mathbf{Y} + \mathbf{b}$. 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 [12] establishes that $\mathbf{A}^*$ and $\mathbf{b}^*$ obey the relation

$$\mathrm{E}_{\mathbf{XY}}\left\{[\mathbf{X}-\mathbf{A}^*\mathbf{Y}-\mathbf{b}^*]\,[\mathbf{A}\mathbf{Y}+\mathbf{b}]^{\mathrm{T}}\right\} = \mathbf{O} \;\; \forall \mathbf{A}, \mathbf{b}. \quad (2)$$

Choosing $\mathbf{A}$ equal to zero, to satisfy the orthogonality condition for all vectors $\mathbf{b}$ it follows that

$$\mathrm{E}_{\mathbf{XY}}\{[\mathbf{X}-\mathbf{A}^*\mathbf{Y}-\mathbf{b}^*]\} = \mathbf{0} \;\Rightarrow\; \mathbf{b}^* = \mathrm{E}\{\mathbf{X}\} - \mathbf{A}^*\,\mathrm{E}\{\mathbf{Y}\}. \quad (3)$$

The linear MMSE (LMMSE) estimator has the form $\hat{\mathbf{X}} = \boldsymbol{\mu}_{\mathbf{X}} + \mathbf{A}^*(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})$.

Choosing $\mathbf{b} = -\boldsymbol{\mu}_{\mathbf{Y}}$ and $\mathbf{A}$ the identity matrix it follows that

$$\mathrm{E}_{\mathbf{XY}}\left\{[\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}}-\mathbf{A}^*(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})]\,(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})^{\mathrm{T}}\right\} = \mathbf{O}, \quad (4)$$

the optimal matrix is therefore given by

$$\mathbf{A}^* = \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}. \quad (5)$$

The LMMSE estimator is

$$\hat{\mathbf{X}} = \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\,(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}}), \quad (6)$$

and has covariance

$$\boldsymbol{\Sigma}_{\hat{\mathbf{X}}\hat{\mathbf{X}}} = \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}^{\mathrm{T}}. \quad (7)$$

The estimation error $\mathbf{E}$ is defined as $\mathbf{E} \triangleq \mathbf{X} - \hat{\mathbf{X}}$. The LMMSE estimator is unbiased, i.e. the estimation error is zero mean. The covariance of the estimation error is

$$\boldsymbol{\Sigma}_{\mathbf{EE}} = \boldsymbol{\Sigma}_{\mathbf{XX}} - \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}^{\mathrm{T}}. \quad (8)$$

Since the estimator is unbiased the error covariance and mean square coincide, therefore LMMSE and linear unbiased minimum variance have the same meaning. The global MMSE estimator (i.e. possibly non-linear) of $\mathbf{X}$ based on $\mathbf{Y}$ is the conditional mean $\mathrm{E}_{\mathbf{X}|\mathbf{Y}}\{\mathbf{X}|\mathbf{Y}\}$.

The UKF approximates $\boldsymbol{\mu}_{\mathbf{Y}}$ and the second order moments in Eqs. (6) and (8) using stochastic linearization through a set of regression points while the EKF approximates them by linearization around the mean.

The following example illustrates the consequences of choosing a linear estimator. Define two independent Gaussian random variables as $Z \sim n(\mu_Z, \sigma^2_{ZZ})$ and $N \sim n(0, \sigma^2_{NN})$. The notation $n(\mu, \sigma^2)$ indicates a Gaussian random variable with mean μ and variance σ^2. The random variable to be estimated is $X = Z + N$, therefore X is Gaussian with mean $\mu_X = \mu_Z$ and variance $\sigma^2_{XX} = \sigma^2_{ZZ} + \sigma^2_{NN}$. Suppose the observed random variable is $Y = (X-N)^3$, therefore the measurement is nonlinear and N is interpreted as measurement noise. Notice that $Y = Z^3$ therefore

$$\begin{aligned}
\mu_Y &= \mathrm{E}\left\{Z^3\right\} = \mu_Z^3 + 3\mu_Z\,\sigma^2_{ZZ} && (9)\\
\sigma^2_{YY} &= \mathrm{E}\left\{Y^2\right\} - \mu_Y^2 = \mathrm{E}\left\{Z^6\right\} - (\mu_Z^3 + 3\mu_Z\,\sigma^2_{ZZ})^2 \\
&= \mu_Z^6 + 15\mu_Z^4\sigma^2_{ZZ} + 45\mu_Z^2\sigma^4_{ZZ} + 15\sigma^6_{ZZ} \\
&\quad - (\mu_Z^6 + 6\mu_Z^4\sigma^2_{YY} + 9\mu_Z^2\sigma^4_{YY}) \\
&= 9\mu_Z^4\sigma^2_{ZZ} + 36\mu_Z^2\sigma^4_{ZZ} + 15\sigma^6_{ZZ} && (10)\\
\sigma^2_{XY} &= \mathrm{E}\{XY\} - \mu_X\mu_Y \\
&= \mathrm{E}\left\{Z^4 + Z^3N\right\} - \mu_Z(\mu_Z^3 + 3\mu_Z\,\sigma^2_{ZZ}) \\
&= \mu_Z^4 + 6\mu_Z^2\sigma^2_{ZZ} + 3\sigma^4_{ZZ} - \mu_Z^4 - 3\mu_Z^2\sigma^2_{ZZ} \\
&= 3\mu_Z^2\sigma^2_{ZZ} + 3\sigma^4_{ZZ}. && (11)
\end{aligned}$$

A possible estimator of X is $\hat{X}_{MEAS} = Y^{1/3}$. Notice that $\hat{X}_{MEAS} = Z$ and has estimation error $E_{MEAS} = X - \hat{X}_{MEAS} = N$. The estimator is unbiased since

$$\mathrm{E}\{E_{MEAS}\} = \mathrm{E}\{N\} = 0. \quad (12)$$

The estimation error variance is given by

$$\sigma^2_{MEAS} = \sigma^2_{NN}. \quad (13)$$

The optimal LMMSE estimator is given by

$$\hat{X}_{LMMSE} = \mu_X - \frac{\sigma^2_{XY}}{\sigma^2_{YY}}(Y - \mu_Y), \quad (14)$$

and has estimation error $E_{LMMSE} \triangleq X - \hat{X}_{LMMSE}$. The estimation error variance (which is equal to the mean square) is

$$\begin{aligned}
\sigma^2_{LMMSE} &= \sigma^2_{XX} - \frac{\sigma^4_{XY}}{\sigma^2_{YY}} \\
&= \sigma^2_{ZZ} + \sigma^2_{NN} - \frac{9(\mu_Z^2 + \sigma^2_{ZZ})^2\,\sigma^4_{ZZ}}{9\mu_Z^4\sigma^2_{ZZ} + 36\mu_Z^2\sigma^4_{ZZ} + 15\sigma^6_{ZZ}} \\
&= \left(1 - \frac{9(\mu_Z^2 + \sigma^2_{ZZ})^2}{9\mu_Z^4 + 36\mu_Z^2\sigma^2_{ZZ} + 15\sigma^4_{ZZ}}\right)\sigma^2_{ZZ} + \sigma^2_{NN} \\
&= \frac{18\mu_Z^2\sigma^2_{ZZ} + 6\sigma^4_{ZZ}}{9\mu_Z^4 + 36\mu_Z^2\sigma^2_{ZZ} + 15\sigma^4_{ZZ}}\sigma^2_{ZZ} + \sigma^2_{NN}. \quad (15)
\end{aligned}$$

In a mean-square sense, the optimal linear estimator performs worse (i.e. it has a bigger mean-square error) than $\hat{X}_{MEAS}$. It is not surprising that a nonlinear estimator performs better than a linear estimator. It is however undesirable that

the linear estimator is not able to extract all the available information from the measurement and performs worse than the measurement itself. A perfectly tuned UKF is consistent and has the same performance of the LMMSE estimator, but this fact does not imply satisfactory results. Notice that to calculate all the variances and cross-variances, the LMMSE estimator necessitates, in general, to know the distribution of the random variables. When the measurements are linear only the knowledge of the first two moments is required.

Because of independence, the joint pdf of Z and N is given by

$$f_{ZN}(z,n) = f_Z(z)\, f_N(n). \tag{16}$$

The transformation between $[Z,N]$ and $[X,Y]$ is

$$\begin{bmatrix} Z \\ N \end{bmatrix} = g^{-1}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = \begin{bmatrix} Y^{1/3} \\ X - Y^{1/3} \end{bmatrix}. \tag{17}$$

The absolute value of the determinant of the Jacobian is given by

$$\left|\frac{\partial g^{-1}}{\partial \begin{bmatrix} X & Y \end{bmatrix}^{\mathrm{T}}}\right| = \left|\begin{bmatrix} 0 & (1/3)Y^{-2/3} \\ 1 & -(1/3)Y^{-2/3} \end{bmatrix}\right| = \frac{1}{3}Y^{-2/3}. \tag{18}$$

The distributions of interest are

$$f_{XY}(x,y) = \frac{1}{3}y^{-2/3} f_Z(y^{1/3})\, f_N(x - y^{1/3}) \tag{19}$$

$$f_Y(y) = \frac{1}{3}y^{-2/3} f_Z(y^{1/3}) \tag{20}$$

$$f_{X|Y}(x|y) = f_N(x - y^{1/3}). \tag{21}$$

The MMSE (non-linear) estimator is

$$\hat{X}_{MMSE} = \mathrm{E}_{X|Y}\{X|Y\} = Y^{1/3} + \mu_N = Y^{1/3}, \tag{22}$$

where $\mu_N = 0$ is used.

III. The Recursive Update Filter

The Kalman filter's linear update is applied all at once when a measurement becomes available. The idea behind the proposed scheme is to apply the update gradually, allowing to recalculate the Jacobian and to "follow" the nonlinearity of the measurement as the update is applied.

Lower case letters indicate realizations of the random variables. For example $\mathbf{x}$ is the true state which amounts to the value of of the random variable $\mathbf{X}$ for the event under consideration.

Recall the Kalman filter with correlated measurement and process noise [13], define the cross-covariance at time t_k between the true (unknown) state $\mathbf{x}_k$ and the zero-mean measurement noise $\boldsymbol{\eta}_k$ as $\mathbf{C}_k$

$$\mathbf{C}_k = \mathrm{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)\boldsymbol{\eta}_k^{\mathrm{T}}\right\}, \tag{23}$$

where $\hat{\mathbf{x}}_k^-$ is the *a priori* estimated state. To derive the equations of this scheme a linear measurement $\mathbf{y}_k$ is first assumed

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\eta}_k. \tag{24}$$

Choosing a linear unbiased update

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-). \tag{25}$$

The optimal gain in terms of minimum variance estimation is given by

$$\mathbf{K}_k = (\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k)\,(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k\mathbf{C}_k + \mathbf{C}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}})^{-1}, \tag{26}$$

where $\mathbf{P}_k^-$ is the *a priori* estimation error covariance matrix and $\mathbf{R}_k$ is the measurement error covariance matrix. It is assumed throughout this work that all errors and noises are zero mean. The updated estimation error covariance is given by

$$\begin{aligned}\mathbf{P}_k^+ =& (\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}} - \\ & (\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)\mathbf{C}_k\mathbf{K}_k^{\mathrm{T}} - \mathbf{K}_k\mathbf{C}_k^{\mathrm{T}}(\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}},\end{aligned} \tag{27}$$

where $\mathbf{I}_{n\times n}$ is the $n \times n$ identity matrix, n being the size of the state vector.

Assume the same measurement is processed twice, after the first update the *a posteriori* estimation error is

$$\mathbf{e}_k^{(1)} = \mathbf{x}_k - \hat{\mathbf{x}}_k^{(1)} = (\mathbf{I}_{n\times n} - \mathbf{K}_k^{(1)}\mathbf{H}_k)\mathbf{e}_k^- - \mathbf{K}_k^{(1)}\boldsymbol{\eta}_k, \tag{28}$$

where $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$ is the *a priori* estimation error. The first optimal gain is

$$\mathbf{K}_k^{(1)} = \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}. \tag{29}$$

The cross-covariance of $\mathbf{e}_k^{(1)}$ and $\boldsymbol{\eta}_k$ is given by

$$\mathbf{C}_k^{(1)} = -\mathbf{K}_k\mathbf{R}_k. \tag{30}$$

The updated covariance is obtained simplifying Eq. (27) to obtain

$$\mathbf{P}_k^{(1)} = (\mathbf{I}_{n\times n} - \mathbf{K}_k^{(1)}\mathbf{H}_k)\mathbf{P}_k^- - \mathbf{K}_k\mathbf{C}_k^{\mathrm{T}}. \tag{31}$$

The cross-covariance between the *a priori* state and the measurement error is denotes as $\mathbf{C}_k^{(0)}$ and is assumed to be zero, i.e. the quantities are uncorrelated and $\mathbf{C}_k^{(0)} = \mathbf{C}_k = \mathbf{O}$. Eq. (31) is only valid when the gain is chosen as the optimal gain, which is not always true in the remaining of this section. Eq. (27) is valid for any choice of $\mathbf{K}_k$.

Processing the same measurement again the second optimal gain is obtained by substituting Eq. (30) into Eq. (26) and replacing $\mathbf{P}_k^-$ with $\mathbf{P}_k^{(1)}$

$$\begin{aligned}\mathbf{K}_k^{(2)} =& (\mathbf{P}_k^{(1)}\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k^{(1)}) \\ & (\mathbf{H}_k\mathbf{P}_k^{(1)} - \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k\mathbf{C}_k^{(1)} + \mathbf{C}_k^{(1)\mathrm{T}}\mathbf{H}_k^{\mathrm{T}})^{-1},\end{aligned} \tag{32}$$

the resulting optimal gain $\mathbf{K}_k^{(2)}$ is zero since

$$\begin{aligned}\mathbf{P}_k^{(1)}\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k^{(1)} &= (\mathbf{I}_{n\times n} - \mathbf{K}_k^{(1)}\mathbf{H}_k)\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \mathbf{K}_k^{(1)}\mathbf{R}_k \\ &= \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \mathbf{K}_k^{(1)}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k) \\ &= \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} = \mathbf{O}.\end{aligned} \tag{33}$$

This result is to be expected; after the first update all the information from the measurement is extracted, therefore processing the same measurement again, no additional update should occur.

Assume however that the first update is not optimal, only a fraction of the optimal update is applied

$$\mathbf{K}_k^{(1)} = 0.5\,\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}, \tag{34}$$

with this choice of $\mathbf{K}_k^{(1)}$ the resulting $\mathbf{K}_k^{(2)}$ is not zero. After the first iteration only half of the optimal update is applied. During the second iteration the full $\mathbf{K}_k^{(2)}$ is applied such that the net result after both updates is identical to the standard Kalman algorithm. If three iterations are performed, and each iteration updates one third of the total, the first coefficient is $1/3$. The second coefficient is $1/2$ because the remaining optimal update is two thirds of the total, the last coefficient is 1. This procedure can be expanded to an arbitrary number N of iterations. For the linear measurement case this algorithm is equivalent to the Kalman filter, making the iterations redundant. The benefits of this approach are evident however for nonlinear measurements. Given a measurement which is a nonlinear function of the state

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \boldsymbol{\eta}_k \tag{35}$$

the algorithm for the nonlinear recursive update is given by Table I.

TABLE I
RECURSIVE UPDATE, NONLINEAR MEASUREMENTS

$\mathbf{C}_k^{(0)} = \mathbf{O},\ \mathbf{P}_k^{(0)} = \mathbf{P}_k^-,\ \hat{\mathbf{x}}_k^{(0)} = \hat{\mathbf{x}}_k^-$
for $i = 1$ to N
$\mathbf{H}_k^{(i)} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{x}}_k^{(i-1)}}$
$\mathbf{W}_k^{(i)} = \mathbf{H}_k^{(i)}\mathbf{P}_k^{(i-1)}\mathbf{H}_k^{(i)\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k^{(i)}\mathbf{C}_k^{(i-1)} + \mathbf{C}_k^{(i-1)\mathrm{T}}\mathbf{H}_k^{(i)\mathrm{T}}$
$\mathbf{K}_k^{(i)} = \gamma_k^{(i)}\big(\mathbf{P}_k^{(i-1)}\mathbf{H}_k^{(i)\mathrm{T}} + \mathbf{C}_k^{(i-1)}\big)\big(\mathbf{W}_k^{(i)}\big)^{-1}$
$\hat{\mathbf{x}}_k^{(i)} = \hat{\mathbf{x}}_k^{(i-1)} + \mathbf{K}_k^{(i)}\big(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^{(i-1)})\big)$
$\mathbf{P}_k^{(i)} = \big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)\mathbf{P}_k^{(i-1)}\big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)^{\mathrm{T}} + \mathbf{K}_k^{(i)}\mathbf{R}_k\mathbf{K}_k^{(i)\mathrm{T}} - \big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)\mathbf{C}_k^{(i)}\mathbf{K}_k^{(i)\mathrm{T}} - \mathbf{K}_k^{(i)}\mathbf{C}_k^{(i)\mathrm{T}}\big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)^{\mathrm{T}}$
$\mathbf{C}_k^{(i)} = \big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)\mathbf{C}_k^{(i-1)} - \mathbf{K}_k^{(i)}\mathbf{R}_k$
$\gamma_k^{(i)} = 1/(N+1-i)$
end for
$\mathbf{P}_k^+ = \mathbf{P}_k^{(N)},\ \hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^{(N)}$

As the number of recursions steps N tends to infinity, it is possible to obtain differential equations that govern an update law that continuously re-linearizes. Define the independent real variable $\tau \in [0,\ 1]$. Assume the domain of τ is divided into N intervals, each of length $1/N$. Then the algorithm of Table I can be rewritten as

$\mathbf{C}_k(0) = \mathbf{O},\ \mathbf{K}_k(0) = \mathbf{O},\ \mathbf{P}_k(0) = \mathbf{P}_k^-,\ \hat{\mathbf{x}}_k(0) = \hat{\mathbf{x}}_k^-$
for $i = 1$ to N
$\tau_i = i/N = \tau_{i-1} + 1/N,\ \tau_0 = 0$
$\mathbf{H}_k(\tau_i) = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{x}}_k(\tau_{i-1})}$
$\mathbf{C}_k(\tau_i) = \big(\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_{i-1})\mathbf{H}_k(\tau_i)\big)\mathbf{C}_k(\tau_{i-1}) - \mathbf{K}_k(\tau_{i-1})\mathbf{R}_k$
$\gamma_k(\tau_i) = (1/N)/(1 + 1/N - \tau_i)$
$\mathbf{W}_k(\tau_i) = \mathbf{H}_k(\tau_i)\mathbf{P}_k(\tau_{i-1})\mathbf{H}_k^{\mathrm{T}}(\tau_i) + \mathbf{R}_k + \mathbf{H}_k(\tau_i)\mathbf{C}_k(\tau_i) + \mathbf{C}_k^{\mathrm{T}}(\tau_i)\mathbf{H}_k^{\mathrm{T}}(\tau_i)$
$\mathbf{K}_k(\tau_i) = \gamma_k(\tau_i)\big(\mathbf{P}_k(\tau_{i-1})\mathbf{H}_k^{\mathrm{T}}(\tau_i) + \mathbf{C}_k(\tau_i)\big)\mathbf{W}_k^{-1}(\tau_i)$
$\hat{\mathbf{x}}_k(\tau_i) = \hat{\mathbf{x}}_k(\tau_{i-1}) + \mathbf{K}_k(\tau_i)\big[\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k(\tau_{i-1}))\big]$
$\mathbf{P}_k(\tau_i) = \big(\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i)\big)\mathbf{P}_k(\tau_{i-1})\big(\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i)\big)^{\mathrm{T}} + \mathbf{K}_k(\tau_i)\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}}(\tau_i) - \big(\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i)\big)\mathbf{C}_k(\tau_i)\mathbf{K}_k^{\mathrm{T}}(\tau_i) - \mathbf{K}_k(\tau_i)\mathbf{C}_k^{\mathrm{T}}(\tau_i)\big(\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i)\big)^{\mathrm{T}}$
end for
$\mathbf{P}_k^+ = \mathbf{P}_k(1),\ \hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k(1)$

Notice that as $N \to \infty$, $\gamma_k \to 0$, therefore $\mathbf{K}_k \to \mathbf{O}$. A new gain $\mathbf{K}_k^*$ is defined as

$$\begin{aligned}\mathbf{K}_k^*(\tau_i) &= \lim_{N\to\infty} \frac{\mathbf{K}_k(\tau_i)}{\Delta\tau} = \lim_{N\to\infty} \frac{\mathbf{K}_k(\tau_i)}{1/N} \\ &= \frac{1}{1-\tau_i}\left(\mathbf{P}_k(\tau_{i-1})\mathbf{H}_k^{\mathrm{T}}(\tau_i) + \mathbf{C}_k(\tau_i)\right)\mathbf{W}_k^{-1}(\tau_i).\end{aligned} \tag{36}$$

The change in cross-covariance between time steps is given by

$$\begin{aligned}\Delta\mathbf{C}_k(\tau_i) &= \mathbf{C}_k(\tau_i) - \mathbf{C}_k(\tau_{i-1}) \\ &= -\mathbf{K}_k(\tau_{i-1})\mathbf{H}_k(\tau_i)\mathbf{C}_k(\tau_{i-1}) - \mathbf{K}_k(\tau_{i-1})\mathbf{R}_k.\end{aligned} \tag{37}$$

Similarly the changes in covariance and state estimate are

$$\Delta\hat{\mathbf{x}}_k(\tau_i) = \mathbf{K}_k(\tau_i)[\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k(\tau_{i-1}))] \tag{38}$$

$$\begin{aligned}\Delta\mathbf{P}_k(\tau_i) &= -\mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i)\mathbf{P}_k(\tau_i) - \mathbf{P}_k(\tau_i)\mathbf{H}_k^{\mathrm{T}}(\tau_i)\mathbf{K}_k^{\mathrm{T}}(\tau_i) \\ &\quad + \mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i)\mathbf{P}_k(\tau_i)\mathbf{H}_k^{\mathrm{T}}(\tau_i)\mathbf{K}_k(\tau_i) \\ &\quad - \mathbf{K}_k(\tau_i)\mathbf{C}_k^{\mathrm{T}}(\tau_i)(\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_i)\mathbf{H}_k^{\mathrm{T}}(\tau_i) \\ &\quad - (\mathbf{I}_{n\times n} - \mathbf{K}_k(\tau_i)\mathbf{H}_k(\tau_i))\mathbf{C}_k(\tau_i)\mathbf{K}_k^{\mathrm{T}}(\tau_i) \\ &\quad + \mathbf{K}_k^{\mathrm{T}}(\tau_i)\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}}(\tau_i)\end{aligned} \tag{39}$$

The evolution of the cross-covariance is given by

$$\begin{aligned}\mathbf{C}_k'(\tau) &= \lim_{N\to\infty} \frac{\mathbf{C}_k(\tau)}{\Delta\tau} \\ &= -\mathbf{K}_k^*(\tau)\mathbf{H}_k(\tau_i)\mathbf{C}_k(\tau) - \mathbf{K}_k^*(\tau)\mathbf{R}_k\end{aligned} \tag{40}$$

Notice that $\mathbf{K}_k^*(1) = \mathbf{O}$ because when $\tau = 1$ all the information from the measurements has been extracted. The evolutions in covariance and state estimate are

$$\hat{\mathbf{x}}_k'(\tau) = \lim_{N\to\infty} \frac{\hat{\mathbf{x}}_k(\tau)}{\Delta\tau} = \mathbf{K}_k^*(\tau)(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k(\tau))] \tag{41}$$

$$\begin{aligned}\mathbf{P}_k'(\tau) &= -\mathbf{K}_k^*(\tau)\left(\mathbf{P}_k(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau_i) + \mathbf{C}_k(\tau)\right) - \\ &\qquad \left(\mathbf{P}_k(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau_i) + \mathbf{C}_k(\tau)\right)^{\mathrm{T}}\mathbf{K}_k^{*\mathrm{T}}(\tau).\end{aligned} \tag{42}$$

For the linear measurement case, integrating Eqs. (40), (41), and (42) from $\tau = 0$ to $\tau = 1$ is equivalent to the standard Kalman update. The differential update algorithm in the presence of nonlinear measurements is given by Table II.

TABLE II
DIFFERENTIAL UPDATE, NONLINEAR MEASUREMENTS

$\mathbf{C}_k(0) = \mathbf{O},\ \mathbf{P}_k(0) = \mathbf{P}_k^-,\ \hat{\mathbf{x}}_k(0) = \hat{\mathbf{x}}_k^-$
Integrate $\mathbf{C}_k$, $\hat{\mathbf{x}}_k$, and $\mathbf{P}_k$ for $\tau = 0$ to 1
$\mathbf{H}_k(\tau) = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{x}}_k(\tau)}$
$\mathbf{W}_k(\tau) = \mathbf{H}_k(\tau)\mathbf{P}_k(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau) + \mathbf{R}_k + \mathbf{H}_k(\tau)\mathbf{C}_k(\tau) + \mathbf{C}_k^{\mathrm{T}}(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau)$
$\mathbf{K}_k^*(\tau) = 1/(1-\tau)\big(\mathbf{P}_k(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau) + \mathbf{C}_k(\tau)\big)\mathbf{W}_k^{-1}(\tau)$
$\mathbf{C}_k'(\tau) = -\mathbf{K}_k^*(\tau)\big(\mathbf{H}_k(\tau)\mathbf{C}_k(\tau) + \mathbf{R}_k\big)$
$\hat{\mathbf{x}}_k'(\tau) = \mathbf{K}_k^*(\tau)\big[\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k(\tau))\big]$
$\mathbf{P}_k'(\tau) = -\mathbf{K}_k^*(\tau)\big(\mathbf{P}_k(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau) + \mathbf{C}_k(\tau)\big) - \big(\mathbf{P}_k(\tau)\mathbf{H}_k^{\mathrm{T}}(\tau) + \mathbf{C}_k(\tau)\big)^{\mathrm{T}}\mathbf{K}_k^{*\mathrm{T}}(\tau)$
end Integrate
$\mathbf{P}_k^+ = \mathbf{P}_k(1),\ \hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k(1).$

In the nonlinear measurement case the total update is different from the EKF and is generally nonlinear. In the differential

formulation of the algorithm the Jacobian is computed continuously, therefore the algorithm "follows" the nonlinearity of the measurement. The EKF assumes a linear update and truncates the Taylor series expansion of the residual to first order, this second assumption can be inadequate in certain situations. The algorithms in Table I and Table II have either an infinitesimal update or an update arbitrarily small, therefore the linearization assumption can always be made valid. The proposed algorithms are not globally optimal, to solve the nonlinear minimum variance problem the distribution of the errors needs to be defined.

In practical implementations the differential update needs to be computed numerically which requires an integration step and inevitably introduces numerical errors. In the author's experience, it is more accurate to use 4 recursions from Table I than a single step fourth order Runge-Kutta propagation from Table II; the two approaches require the same number of function evaluations. Using N recursions is equivalent to solving the differential update with N Euler steps. Since $\mathbf{K}_k(1) = \mathbf{O}$ numerical integration schemes that do not evaluate the derivative at the end of the step seem to be more precise.

The number of iterations is a user-defined parameter that needs to be selected by addressing two conflicting design objectives. On the one hand the number of iterations should be chosen high to improve performance; the more iterations the more often the algorithm re-linearizes and the better the non-linearity of the measurement is followed. On the other hand whenever computational time is of concern it is desirable to reduce the number of iterations. In general, the higher the degree of nonlinearity of the measurement, the more iterations are needed. A good indicator of the performance of the algorithm is the post-update residual. The residual (actual measurement minus estimated measurement) where the estimated measurement is computed with the updated state, should match its predicted covariance. Discrepancies between the two indicate the nonlinear effects are of concern and more iterations are needed.

IV. Comparison with Existing Schemes

The algorithms proposed in this paper have two characteristics that differentiates them from other schemes. The EKF can suffer divergence when nonlinearities become significant [9]. For example, it is known that the linearization assumption of the EKF is not adequate in the presence of very accurate nonlinear measurement when the prior estimation error covariance is large [14]. In this circumstance the update is "large" and the actual covariance often does not decrease as fast as the filter's predicted covariance. The proposed algorithms mitigate this issue by substituting a single "large" update with small or infinitesimal updates. Algorithms such the GSOF and the UKF are able to capture some of the nonlinearities and are more robust to these divergence problems than the EKF.

The second feature of the proposed schemes is the nonlinearity of the total update. Both the UKF and GSOF have a linear update. Since the update is not applied all at once and the Jacobian is re-calculated, the total update is not linear. It is well known that in the presence of nonlinear measurements the linear update is not necessarily optimal.

As seen in the appendix, the Gaussian second order filter mostly consists in increasing the process noise covariance $\mathbf{R}_k$ by the second order contributions $\mathbf{B}_k$. In practice the nonlinearities are treated as additional sensor noise. The posterior uncertainty of the EKF's estimated measurements tends to be equal to the measurement covariance when the prior uncertainty is high and the measurement is precise. Under the same conditions the post-update estimated measurement of the Gaussian second order filter tends to $\mathbf{R}_k + \mathbf{B}_k$. Therefore after an update the second order filter could perform worse than the measurement it is provided in terms of the size of the estimation error covariance. This fact illustrates a possible shortcoming of the linear update in the presence of nonlinear measurements. After an update an estimation algorithm should perform at least as accurately as the measurement it is provided, otherwise the measurement itself would be a better estimate. The UKF behaves in a very similar manner. A quick review of the UKF is also presented in the appendix.

The existing algorithm that most resembles the proposed scheme is the iterated Kalman filter (IKF). The purpose of the iterated Kalman filter update [5] is to repeatedly calculate the measurement Jacobian each time linearizing about the most recent estimate. The iteration is initialized by choosing

$$\hat{\mathbf{x}}_{k,0} = \hat{\mathbf{x}}_k^-, \qquad \mathbf{P}_{k,0} = \mathbf{P}_k^-. \tag{43}$$

The loop is given by

$$\mathbf{K}_{k,i+1} = \mathbf{P}_k^- \mathbf{H}_{k,i}^{\mathrm{T}} \left[\mathbf{H}_{k,i} \mathbf{P}_k^- \mathbf{H}_{k,i}^{\mathrm{T}} + \mathbf{R}_k \right]^{-1} \tag{44}$$

$$\hat{\mathbf{x}}_{k,i+1} = \hat{\mathbf{x}}_k^- + \mathbf{K}_{i,k} \left[\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k,i}) - \mathbf{H}_{k,i}(\hat{\mathbf{x}}_{k,i}) \left(\hat{\mathbf{x}}_k^- - \hat{\mathbf{x}}_{k,i} \right) \right] \tag{45}$$

$$\mathbf{P}_{k,i+1} = \left[\mathbf{I} - \mathbf{K}_{k,i} \mathbf{H}_{k,i} \right] \mathbf{P}_k^- \tag{46}$$

$$\mathbf{H}_{k,i} = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\mathbf{x} = \hat{\mathbf{x}}_{k,i}}. \tag{47}$$

The updated state estimate and covariance are given by

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_{k,N}, \qquad \mathbf{P}_k^+ = \mathbf{P}_{k,N}, \tag{48}$$

where N is the number of iterations. The IKF is fundamentally different from the proposed algorithm because it recursively recalculates the center of the Taylor series expansions and reapplies the entire update to both the state and the covariance. The estimate $\hat{\mathbf{x}}$ provided by the IKF is equivalent to using a Gauss-Newton method [15] to minimize the following nonlinear least-squares problem

$$\begin{aligned} \min_{\mathbf{x}} \mathcal{J} =& (\mathbf{x} - \hat{\mathbf{x}}^-)^{\mathrm{T}} (\mathbf{P}^-)^{-1} (\mathbf{x} - \hat{\mathbf{x}}^-) \\ &+ (\mathbf{h}(\mathbf{x}) - \mathbf{y})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{h}(\mathbf{x}) - \mathbf{y}). \end{aligned} \tag{49}$$

Newton methods converge on an interval when the derivative of the function is non-zero in the interval and the interval is "small" enough, i.e. the higher order effects are not dominating. When the second of these two conditions does not apply the method can overshoot and diverge. Much like the IKF, the proposed recursive update method fails when the derivative is zero (as does the EKF). However, because the update is applied gradually, no overshooting occurs. A very simple example

illustrates this concept. The arctangent function is monotonically increasing and is a well-behaved function. Assume a true state $x = 0$ and a perfect measurement $y = \arctan x$. The measurement sensitivity is $H = \partial y/\partial x = 1/(1 + x^2)$ and is not zero anywhere close to the true state. Since the measurement is perfect $R = 0$ and the Kalman gain is $K = P^- H/(P^- H^2) = 1/H$. Choosing the *a priori* estimate $\hat{x}^- = \hat{x}^{(0)} = 1.5$ the IKF diverges since

$$\hat{x}^{(1)} = -1.694 \qquad \hat{x}^{(2)} = 2.321$$
$$\hat{x}^{(3)} = -5.114 \qquad \hat{x}^{(4)} = 32.295.$$

Under the same conditions, four recursions of the proposed algorithm provide

$$\hat{x}^{(1)} = 0.701 \qquad \hat{x}^{(2)} = 0.397$$
$$\hat{x}^{(3)} = 0.178 \qquad \hat{x}^{(4)} = -0.004.$$

For this particular case, since the measurement is perfect, the minimum of Eq. (49) is zero and the IKF reduces to a Newton-Raphson method.

As long as the measurement sensitivity does not approach zero, in most practical situations both the IKF and the recursive update filter are equally adequate to solve this type of problems. The IKF requires less computations, but the proposed method is preferable under certain conditions like the one illustrated above. While the IKF is a Gauss-Newton method, the recursive update filter closely resembles a gradient method. Line searching techniques [16] can avoid divergence of Newton methods, they usually rely on numerical optimization and their complexity is not usually suitable for real-time applications.

V. Numerical Examples

In order to demonstrate the advantages of the proposed algorithm two examples are presented here. The first is a simple scalar example that provides a good geometrical interpretation of the difference between the various estimation techniques. Assume the true state is $x = 3.5$, the *a priori* estimated state is $\hat{x}^- = 2.5$. The *a priori* estimation error covariance is $P^- = .5^2$ such that the *a priori* error falls exactly at the 2-sigma point. The measurement is nonlinear and given by

$$y = x^3 + \eta, \tag{50}$$

with $\eta \sim n(0, 0.1^2)$, therefore the measurement is very accurate. A possible estimate is obtained from the measurement only and is given by

$$\hat{x}_{MEAS} = y^{1/3}. \tag{51}$$

A simple Monte Carlo simulation shows that for this case where $R = 0.1^2$, the estimation error covariance of $\hat{x}_{MEAS}$ is given by $P_{MEAS} = 0.0027^2$, which is very close to its linear approximation: the measurement Jacobian is $H = 3x^2$ and $P_{MEAS} \simeq R/H^2$. By evaluating H at the true state $P_{MEAS} \simeq .1^2/(3 \cdot 3.5^2)^2 = 0.0027^2$.

The EKF gain is given by

$$K_{EKF} = P^-[3(\hat{x}^-)^2]/[9(\hat{x}^-)^4 P^- + R] = 0.0533. \tag{52}$$

To keep the analysis deterministic we assume that the particular measurement under consideration happens to be error free, i.e. $y = 42.875$. The updated state is equal to

$$\hat{x}^+_{EKF} = \hat{x}^- + K_{EKF}\,[y - (\hat{x}^-)^3] = 3.9532, \tag{53}$$

the estimation error is equal to 0.4532. The *a posteriori* error covariance is

$$P^+_{EKF} = [1 - 3K_{EKF}(\hat{x}^-)^2]^2 P^- + K^2_{EKF} R = 0.0053^2. \tag{54}$$

The estimation error is 85 times its expected standard deviation. Since the measurement is much more accurate than the *a priori* estimate, the *a posteriori* estimate is mainly based on the measurement. The EKF "thinks" it extracts all the information but it has two issues. First the *a posteriori* covariance is calculated based on the *a priori* estimate rather than the true state, $P^+_{EKF} \simeq R/[3(\hat{x}^-)^2]^2 = 0.0053^2$. The second problem is that the filter is not actually able to extract all the information from the measurement because it performs a linear update. Fig. 1 gives a graphical representation of the EKF update. The x-axis is the state while the y-axis is the measurement, the gray line is $y = x^3$. The EKF correctly recognizes that the measurement is much more precise than the *a priori* information, therefore it attempts to update the state to match the measurement. However, relying on linearization, the EKF does not match $(\hat{x}^+)^3$ with the measurement, instead it matches its first order approximation

$$(\hat{x}^+)^3 \simeq \hat{y}^- + H\,K_{EKF}(\hat{x}^+ - \hat{x}^-) = (\hat{x}^-)^3 + 3(\hat{x}^-)^2 \Delta\hat{x}, \tag{55}$$

in this situation the first order approximation is not accurate. Fig. 1 shows that the approximated *a posteriori* measurement estimate matches very closely the actual measurement, unfortunately in this case it does not imply good performance.

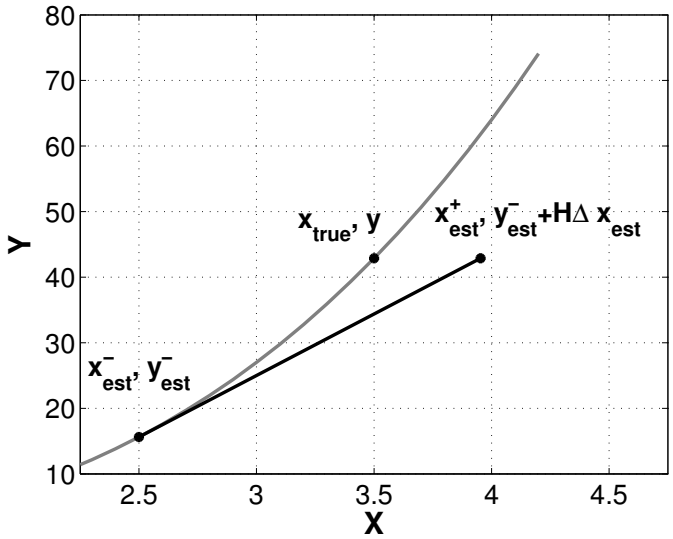

Fig. 1. EKF Update.

The Gaussian second order filter improves the estimate a little, but its biggest effect is to increase the *a posteriori* error covariance. The principal merit of the GSOF is to represent

the fact that the update is not as good as it should be, but it does not actually improve it. The Hessian of the measurement is given by $H' = 6x$. The GSOF gain is given by

$$K_{GSOF} = P(3\hat{x}^2)/[9\hat{x}^4 P + R + 0.5(6\hat{x}^-)^2(P^-)^2] = 0.0494, \tag{56}$$

the gain is only slightly smaller that the EKF's gain. The updated state is

$$\hat{x}^+_{GSOF} = \hat{x}^- + K_{GSOF}\,[y - (\hat{x}^-)^3 - 0.5(6\hat{x}^-)P^-] = 3.7530, \tag{57}$$

the estimation error is equal to 0.253. The *a posteriori* error covariance is

$$\begin{aligned} P^+_{GSOF} =& [1 - 3K_{GSOF}(\hat{x}^-)^2]^2 \hat{P}^- \\ &+ K^2_{GSOF}[R + 0.5(6\hat{x}^-)^2(P^-)^2] = 0.1362^2, \end{aligned} \tag{58}$$

the error estimation error is 1.86 times its expected standard deviation. The filter is able to correctly predict the estimation error, but it mainly accomplishes that by increasing the uncertainty rather than improving the estimate.

Since x and η are Gaussian, the UKF parameter is chosen as $\kappa = 2$ [10]. For this particular case, the UKF outputs are

$$K_{UKF} = 0.0513 \tag{59}$$

$$\hat{x}^+_{UKF} = 3.8654 \tag{60}$$

$$P^+_{UKF} = 0.1688 = 0.4109^2. \tag{61}$$

The UKF has the biggest estimated covariance and the estimation error is 0.89 times the expected standard deviation. The UKF achieves consistency and minimizes the mean square error out of all linear updates. The performance however is not ideal because it is worse than the measurement accuracy. The discussion in section II highlights that a nonlinear update is necessary to overcome this issue.

Using 10 steps in the recursive update filter the results are

$$\hat{x}^+_{RUF} = 3.5014 \tag{62}$$

$$P^+_{RUF} = 8.0234\ 10^{-6} = 0.0028^2. \tag{63}$$

The ratio of the estimated error over its expected standard deviation is 0.4993. Fig. 2 and Fig. 3 show the performance of RUF with 2 and 10 steps, respectively.

To further illustrate the difference between the proposed algorithm and the IKF, two iterations are used and Fig. 2 is compared with Fig. 4. Two recursions in the proposed algorithm result in an estimate of $\hat{x}^{(2)}_{RUF} = 3.5238$ while two iterations IKF produce $\hat{x}^{(2)}_{IKF} = 3.5499$. Under the conditions of this example the proposed algorithm outperforms the IKF. As the number of iterations increase the IKF performance improves significantly and the IKF converges fast. Fig. 4 shows how the IKF operates. The true state is at center of the plot, while the initial estimate is at the bottom left. The first IKF iteration is identical to the EKF and happens along the tangent of the measurement function at the *a priori* estimation point. The IKF then recalculates the tangent of the measurement function at the current estimation point (thinner dashed line in the figure). The second IKF iteration applies the entire update along the recalculated tangent starting from the *a priori* estimation point.

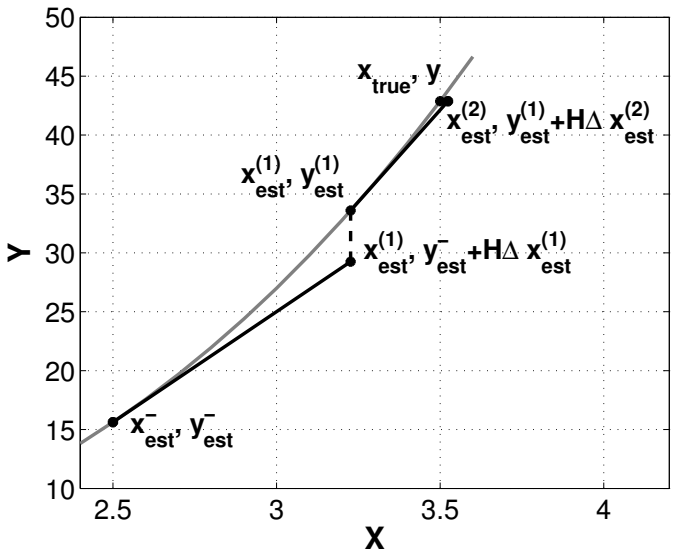

Fig. 2. RUF Update with 2 steps.

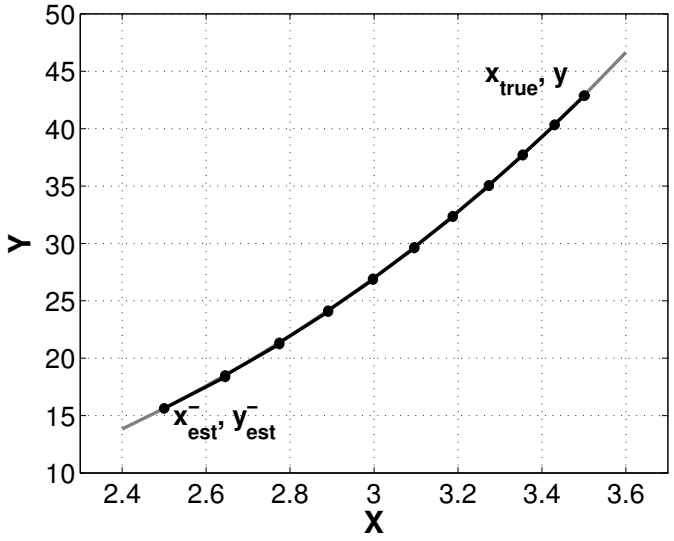

Fig. 3. RUF Update with 10 steps.

Fig. 5 shows the algorithms comparison for 100 Monte Carlo runs. For each algorithm the 100 estimation errors are plotted together with the filter's 3σ predictions. All 100 runs have the same *a priori* estimate, while the true state is randomized according to a normal distribution with mean $\hat{x}^-$ and variance P^-. The measurement error is also randomized with mean zero and variance R. The results show that the EKF is unable to predict its performance. The GSOF performs better, but still underestimates the estimation error, probably higher order filters are needed. The UKF is able to correctly predict its performance. Ten iterations are chosen for both the recursive update filter and the IKF. Notice that under the conditions of this example, RUF and IKF are the only algorithms for which the *a posteriori* covariance varies for each Monte Carlo run. In Fig. 5 the covariance of the measurement-only update is the

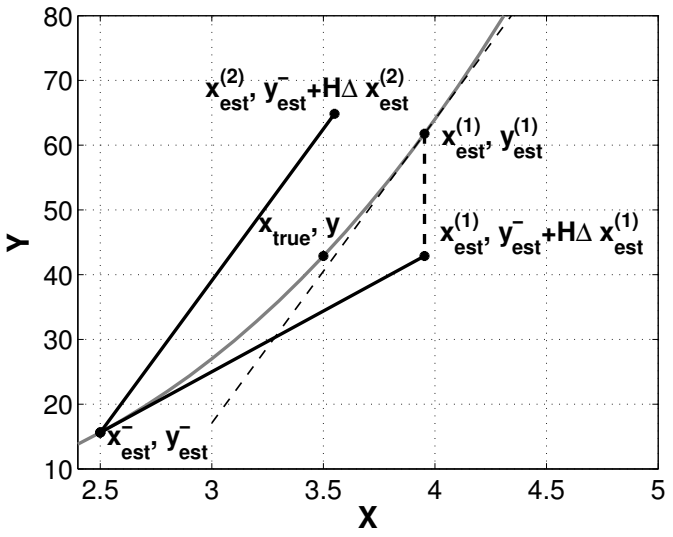

Fig. 4. IKF Update.

sample covariance of 10000 Monte Carlo runs. RUF and IKF reflect the fact that the bigger the true state, the smaller the covariance ($P^+ \simeq R/H^2$), where the measurement Jacobian is evaluated at the true state).

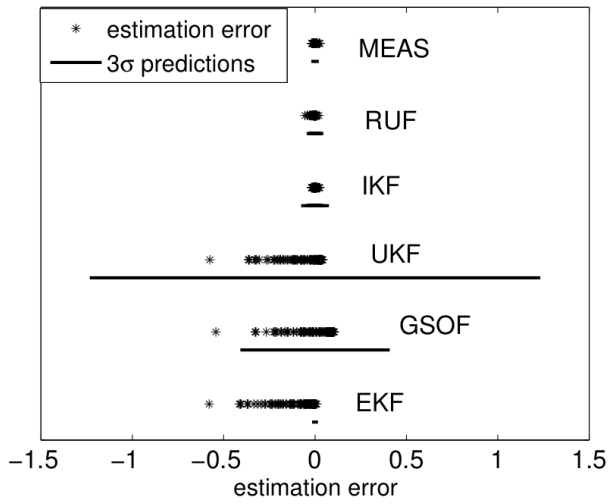

Fig. 5. Estimators Comparison.

The second example presented here is a tracking problem where an object being tracked travels on a planar trajectory as shown in Fig. 6. The estimated state $\mathbf{x}$ contains position and velocity, the initial filter covariance $\mathbf{P}_0$ is obtained setting the state uncertainty to 10 m in position and 0.05 m/s in velocity per axis.

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \boldsymbol{\nu} \qquad \mathbf{P}_0 = \begin{bmatrix} 10^2\mathbf{I}_{3\times3} & \mathbf{O}_{3\times3} \\ \mathbf{O}_{3\times3} & 0.05^2\mathbf{I}_{3\times3} \end{bmatrix} \tag{64}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{O}_{3\times3} & \mathbf{I}_{3\times3} \\ \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \tag{65}$$

and

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -n^2 & 0 \\ 0 & 0 & 3n^2 \end{bmatrix} \qquad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 2n \\ 0 & 0 & 0 \\ -2n & 0 & 0 \end{bmatrix},$$

with $n = 0.0011$rad/s. The process noise is given by

$$\mathrm{E}\{\boldsymbol{\nu}\} = \mathbf{0} \qquad \mathrm{E}\{\boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{T}}\} = \begin{bmatrix} \mathbf{O}_{3\times3} & \mathbf{O}_{3\times3} \\ \mathbf{O}_{3\times3} & 10^{-9}\mathbf{I}_{3\times3} \end{bmatrix} \tag{66}$$

A range measurement ρ with 0.1 m accuracy and two bearing angles with 0.1 deg accuracy are available. The bearing angles are azimuth $\alpha = \tan^{-1}(\mathbf{x}(1)/\mathbf{x}(2))$ and elevation $\epsilon = \sin^{-1}(\mathbf{x}(3)/\rho)$.

$$\mathbf{y} = \begin{bmatrix} \rho \\ \alpha \\ \epsilon \end{bmatrix} + \boldsymbol{\eta} \tag{67}$$

$$\mathrm{E}\{\boldsymbol{\eta}\} = \mathbf{0}, \qquad \mathrm{E}\{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}\} = \begin{bmatrix} 0.1^2 & \mathbf{O}_{1\times2} \\ \mathbf{O}_{2\times1} & (0.1\pi/180)^2\mathbf{I}_{2\times2} \end{bmatrix} \tag{68}$$

One hundred runs are performed in which the initial estimation error and the sensors errors are dispersed using zero mean Gaussian independent random variables. Fig. 7 shows the performance of the EKF, the lighter lines are the 100 samples of the estimation error, while the thicker black lines are the EKF predicted standard deviations. Much like in the previous example the EKF is overly optimistic in predicting its performance. Eventually however the EKF is able to recover. Under similar but more severe circumstances, it is also possible that the EKF would diverge all together [11].

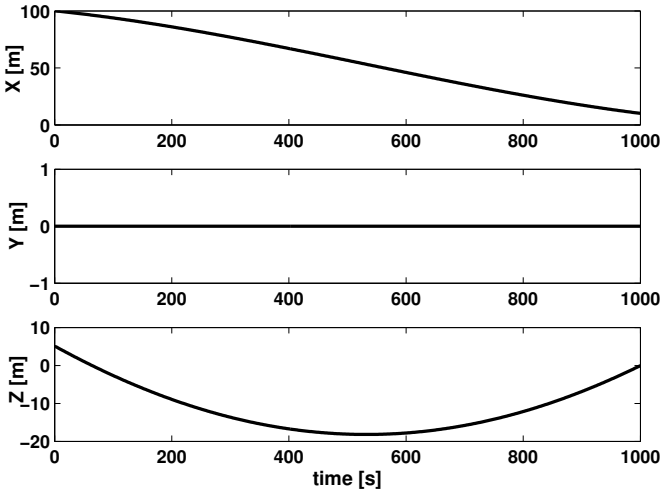

Fig. 6. Trajectory.

Fig. 8 shows the performance of 10 recursive updates using the proposed algorithms. It can be seen that the filter outperforms the EKF during the first few hundred seconds. The recursive update filter is also able to predict its performance as most error samples are within the 1σ prediction and all of them are within the 3σ values. The performance of the two algorithms in estimating velocity is very similar to that of the position shown in Figs. 7 and 8.

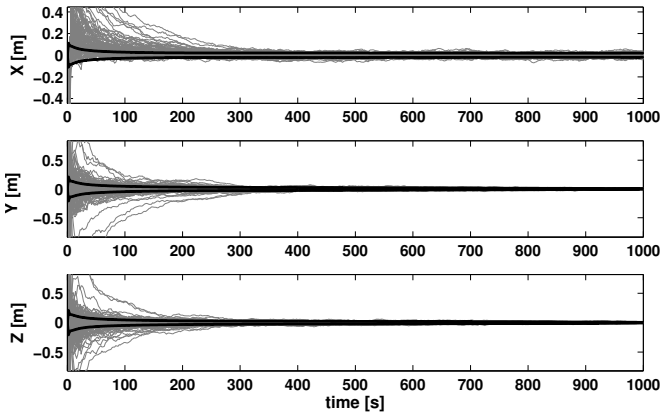

Fig. 7. EKF Estimation Error and Predicted Covariance.

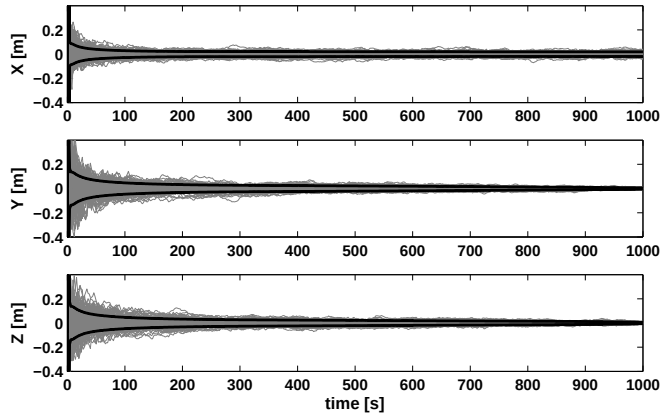

Fig. 8. Recursive Update Filter Estimation Error and Predicted Covariance.

VI. Conclusion

Optimal does not mean good. The linear minimum mean square error (LMMSE) estimator is optimal out of linear estimators, however the feature does not guarantee good performance for certain measurements nonlinearities. In a minimum mean square sense, the LMMSE estimate is always better than the *a priori* estimate (i.e. the estimate without incorporating the measurements) but there is no guarantee it performs better than the measurements themselves. This work proposes a nonlinear estimator that outperforms the extended Kalman filter (EKF) in the presence of non-linear measurements. Like the EKF, the proposed algorithm linearizes the measurement residual but the update is applied a little at the time to assure the validity of the linearization assumption. Like most practical non-linear filters, the proposed algorithm makes simplifications and approximations and it could be outperformed under certain conditions. The main assumption is that the optimal gain follows the measurement gradient. This is often a good assumption and it is an improvement over the EKF. However under some conditions, namely the slope of the measurement function approaches zero, the approach can be unsatisfactory. The EKF also underperforms in these circumstances. In this work the Kalman gain and the measurement Jacobian are taken outside the expected value sign when calculating the filter's covariance matrix as done by the EKF.

Under the conditions of this work's numerical example, the proposed algorithm outperforms the EKF, the Gaussian second order filter, and the unscented Kalman filter. In the example, the proposed algorithm is able to extract all the information from the measurements therefore it has a higher convergence rate than linear estimators such as the UKF or the GSOF. Because of this property the algorithm is particularly useful in highly uncertain environments subject to scarce nonlinear measurements.

Appendix A
The Gaussian Second Order Filter

Given the nonlinear measurement of Eq. (35), the GSOF assumes a linear update

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\big(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k) - \hat{\mathbf{b}}_k\big). \tag{69}$$

The EKF truncates the Taylor series expansion of the nonlinear measurement to first order, while the GSOF includes an additional term, $\hat{\mathbf{b}}_k$, which is the expected value of the second order terms of the Taylor series. The Hessian of the measurement function is given by

$$\mathbf{H}'_{i,k} \triangleq \left[\left. \frac{\partial^2 h_i(\mathbf{x}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^{\mathrm{T}}} \right|_{\mathbf{x}_k = \hat{\mathbf{x}}_k^-} \right],$$

where $h_i(\mathbf{x}_k)$ is the i^{th} component of $\mathbf{h}(\mathbf{x}_k)$. The i^{th} component of $\hat{\mathbf{b}}_k$ is given by

$$\hat{b}_{i,k} = \frac{1}{2}\,\mathrm{trace}(\mathbf{H}'_{i,k}(\mathbf{P}_k^-)^{\mathrm{T}}). \tag{70}$$

Matrix $\mathbf{B}_k$ is the contribution of the second order effects, under the Gaussian approximation, the ij^{th} component of $\mathbf{B}_k$ is given by

$$B_{ij,\,k} = \frac{1}{2}\,\mathrm{trace}(\mathbf{H}'_j(t_k)\,\mathbf{P}_k^-\,\mathbf{H}'_i(t_k)\,\mathbf{P}_k^-) \tag{71}$$

and the posterior estimation error covariance is given by

$$\begin{aligned}\mathbf{P}_k^+ = &(\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_k^-\,(\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} \\ &+ \mathbf{K}_k(\mathbf{R}_k + \mathbf{B}_k)\mathbf{K}_k^{\mathrm{T}}\end{aligned} \tag{72}$$

finally the optimal gain is given by

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{B}_k)^{-1}. \tag{73}$$

In the second order filter the process noise covariance $\mathbf{R}_k$ is increased by the second order contributions $\mathbf{B}_k$.

Appendix B
The Unscented Kalman Filter

The unscented Kalman filter (UKF) [10] is developed assuming that it is more accurate to select a fixed number of points (called sigma points) to approximate a nonlinear transformation of a distribution than to approximate it by linearizing the transformation. This assumption is certainly true for Gaussian distributions, for which optimal locations of the sigma points are known. For general, non-Gaussian, distributions tuning and engineering judgement is needed to

appropriately select the sigma points. Even when the distribution is initially Gaussian, it generally looses Gaussianity after non-linear transformations.

When the measurement noise is additive and uncorrelated, the set of $2n+1$ sigma points $\boldsymbol{\mathcal{X}}_i$ is chosen as

$$\boldsymbol{\mathcal{X}}_0 = \hat{\mathbf{x}}^- \tag{74}$$

$$\boldsymbol{\mathcal{X}}_i = \hat{\mathbf{x}}^- + \left(\sqrt{(n+\kappa)\mathbf{P}^-}\right)_i \tag{75}$$

$$\boldsymbol{\mathcal{X}}_{i+n} = \hat{\mathbf{x}}^- - \left(\sqrt{(n+\kappa)\mathbf{P}^-}\right)_i \tag{76}$$

where n is the size of the state vector, $\left(\sqrt{(n+\kappa)\mathbf{P}^-}\right)_i$ is the i-th column of the matrix square root of $(n+\kappa)\mathbf{P}^-$, and κ is a tuning parameter. The non-additive and/or correlated measurement noise case requires only slight modifications to the algorithm. The set of measurement sigma points $\boldsymbol{\mathcal{Y}}_i$ is calculated as

$$\boldsymbol{\mathcal{Y}}_i = \mathbf{h}(\boldsymbol{\mathcal{X}}_i), \tag{77}$$

the predicted measurement is given by

$$\hat{\mathbf{y}} = \sum_{i=0}^{2n} W_i \boldsymbol{\mathcal{Y}}_i, \tag{78}$$

where the weight is calculated as $W_i = 0.5/(n+\kappa)$ for $i \neq 0$ and $W_0 = \kappa/(n+\kappa)$. The UKF uses Eqs. (6) and (8) to calculate the estimated state and covariance with $\boldsymbol{\mu}_{\mathbf{X}} = \hat{\mathbf{x}}^-$ and $\boldsymbol{\mu}_{\mathbf{Y}} = \hat{\mathbf{y}}$. The UKF approximates the second order moments in those equations as

$$\boldsymbol{\Sigma}_{\mathbf{YY}} = \mathbf{R} + \sum_{i=0}^{2n} W_i \left(\boldsymbol{\mathcal{Y}}_i - \hat{\mathbf{y}}\right)\left(\boldsymbol{\mathcal{Y}}_i - \hat{\mathbf{y}}\right)^{\mathrm{T}} \tag{79}$$

$$\boldsymbol{\Sigma}_{\mathbf{XY}} = \sum_{i=0}^{2n} W_i \left(\boldsymbol{\mathcal{X}}_i - \hat{\mathbf{x}}^-\right)\left(\boldsymbol{\mathcal{Y}}_i - \hat{\mathbf{y}}\right)^{\mathrm{T}}. \tag{80}$$

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, Series D, pp. 35–45, March 1960.

[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction," *Journal of Basic Engineering*, vol. 83, Series D, pp. 95–108, March 1961.

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

[4] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, vol. 7, no. 4, pp. 465–479, July 1971.

[5] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: The MIT press, 1996, pp. 182–189 and pp. 190–191.

[6] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Sciences and Engineering. New York, New York 10003: Academic Press, 1970, vol. 64, pp. 344–346.

[7] D. Alspach and H. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972.

[8] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 2, pp. 107–113, April 1993.

[9] P. S. Mayback, *Stochastic Models, Estimation, And Control Volume 2*, ser. Mathematics in Science and Engineering. New York, NY: Academic Press, 1982.

[10] 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.

[11] R. Zanetti, K. J. DeMars, and R. H. Bishop, "Underweighting Nonlinear Measurements," *Journal of Guidance, Control, and Dynamics*, vol. 33, no. 5, pp. 1670–1675, September–October 2010.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 1st ed. McGraw-Hill, 1965, Secs. 7-4 and 8-2.

[13] R. G. Brown and P. Y. Hwang, *Introduction To Random Signals And Applied Kalman Filtering*, 3rd ed. John Wiley and Sons, 1997, pp. 348–350.

[14] P. J. Huxel and R. H. Bishop, "Fusing Inertial and Relative Measurements for Inertial Navigation in the Presence of Large State Error Covariances," in *Proceedings of the 16th AAS/AIAA Spaceflight Mechanics Meeting*, January 2006.

[15] B. M. Bell and F. W. Cathey, "The Iterated Kalman FIlter Update as a Gauss-Newton Method," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, February 1993.

[16] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, ser. Computational Mathematics. Englewood Cliffs, NJ: Prentice-Hall, 1983.

PLACE PHOTO HERE

Renato Zanetti received his undergraduate degree from Politecnico di Milano, Italy in 2003 and his Ph.D. from the University of Texas at Austin in 2007, both in aerospace engineering.

Since 2007 he has been a senior member of the technical staff at the Charles Stark Draper Laboratory, in the Houston, Texas office. He has been mainly working autonomous navigation for various NASA and private vehicles. His research work focuses on nonlinear/non-Gaussian estimation, attitude dynamics and control, and autonomous vehicles.