AAS 14-345

ADAPTABLE ITERATIVE AND RECURSIVE KALMAN FILTER SCHEMES

Renato Zanetti[*]

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. The Iterated Kalman filter (IKF) and the Recursive Update Filter (RUF) are two algorithms that reduce the consequences of the linearization assumption of the EKF by performing N updates for each new measurement, where N is the number of recursions, a tuning parameter. This paper introduces an adaptable RUF algorithm to calculate N on the go, a similar technique can be used for the IKF as well.

INTRODUCTION

The 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. A widely used algorithm in real-time nonlinear estimation is the extended Kalman filter [3] (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 [4], 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 [5, 6]. Techniques exist to overcome some of the limitations of the EKF linearization assumption. The Gaussian second order filter (GSOF) [7] takes into account second-order terms assuming the error distribution is Gaussian. The iterated extended Kalman filter (IEKF) [3] recursively improves the center of the Taylor series expansion for a better linearization. The unscented Kalman filter [8] (UKF) is able to retain higher-order terms of the Taylor series expansion. Underweighting [9] is an *ad hoc* technique to compensate for the second order effects without actually computing them. The Recursive Update Filter (RUF) [10] applies the update gradually and re-linearizes at each recursion, hence avoiding linearization problems.

The number of iterations in the IKF and RUF are user-defined parameters that needs to be selected by addressing two conflicting design objectives. On the one hand the number of iterations should be

[*] GN&C Engineer, Autonomous Flight Systems, NASA Johnson Space Center, 2101 NASA Parkway, Houston, Texas, 77058.

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 (actual measurement minus estimated measurement computed with the updated state). The residual should match its predicted covariance. Discrepancies between the two indicate the nonlinear effects are of concern and more iterations are needed.

The estimate $\hat{\mathbf{x}}$ provided by the IKF is equivalent to using a Gauss-Newton method [11] to minimize the following nonlinear least-squares problem:

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

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 RUF fails when the derivative is zero (as does the EKF). However, because the update is applied gradually, no overshooting occurs. While the IKF is a Gauss-Newton method, the RUF closely resembles a gradient method. Line searching techniques [12] can avoid divergence of Newton methods, they usually rely on numerical optimization and their complexity is not usually suitable for real-time applications. Standard numerical optimization techniques can be used to calculate the number of iterations or recursions, but they require to repeatedly calculate the performance index Eq. (1), which requires the inversion of the filter covariance matrix, which can prove computationally expensive.

This paper introduces a novel algorithm to adaptably and autonomously select the number of recursions on the go. The algorithm takes advantage of the aforementioned relation between the measurement residuals and their covariance. Only the covariance of the residual needs to be inverted, which is usually of a much smaller dimension than the state covariance.

THE ITERATED KALMAN FILTER

The purpose of the iterated Kalman filter update [3] is to repeatedly calculate the measurement Jacobian each time linearizing about the most recent estimate. The iteration is initialized by choosing

$$\hat{\mathbf{x}}_{k,0}=\hat{\mathbf{x}}_k^-, \qquad\qquad \mathbf{P}_{k,0}=\mathbf{P}_k^-. \tag{2}$$

The loop is given by

$$\mathbf{H}_{k,i}=\left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k,i}} \tag{3}$$

$$\mathbf{K}_{k,i}=\mathbf{P}_k^-\mathbf{H}_{k,i}^{\mathrm{T}}\left[\mathbf{H}_{k,i}\mathbf{P}_k^-\mathbf{H}_{k,i}^{\mathrm{T}}+\mathbf{R}_k\right]^{-1} \tag{4}$$

$$\hat{\mathbf{x}}_{k,i+1}=\hat{\mathbf{x}}_k^-+\mathbf{K}_{i,k}\left[\mathbf{y}_k-\mathbf{h}(\hat{\mathbf{x}}_{k,i})-\mathbf{H}_{k,i}(\hat{\mathbf{x}}_{k,i})\left(\hat{\mathbf{x}}_k^--\hat{\mathbf{x}}_{k,i}\right)\right] \tag{5}$$

$$\mathbf{P}_{k,i+1}=\left[\mathbf{I}-\mathbf{K}_{k,i}\mathbf{H}_{k,i}\right]\mathbf{P}_k^- \tag{6}$$

The updated state estimate and covariance are given by

$$\hat{\mathbf{x}}_k^+=\hat{\mathbf{x}}_{k,N}, \qquad\qquad \mathbf{P}_k^+=\mathbf{P}_{k,N}, \tag{7}$$

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 re-applies the entire update to both the state and the covariance. Notice that both the covariance and the state update, at each iteration always "start" from the prior values $\mathbf{P}_k^-$ and $\hat{\mathbf{x}}_k^-$, this is in contrast with the following, incorrect implementation [13]

$$\hat{\mathbf{x}}_{k,0} = \hat{\mathbf{x}}_k^-, \qquad \mathbf{P}_{k,0} = \mathbf{P}_k^- \tag{8}$$

$$\mathbf{K}_{k,i} = \mathbf{P}_{k,i}\mathbf{H}_{k,i}^{\mathrm{T}}\left[\mathbf{H}_{k,i}\mathbf{P}_{k,i}\mathbf{H}_{k,i}^{\mathrm{T}} + \mathbf{R}_k\right]^{-1} \tag{9}$$

$$\hat{\mathbf{x}}_{k,i+1} = \hat{\mathbf{x}}_{k,i} + \mathbf{K}_{i,k}\left[\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k,i})\right] \tag{10}$$

$$\mathbf{P}_{k,i+1} = \left[\mathbf{I} - \mathbf{K}_{k,i}\mathbf{H}_{k,i}\right]\mathbf{P}_{k,i} \tag{11}$$

$$\mathbf{H}_{k,i} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k,i}} \tag{12}$$

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_{k,N}, \qquad \mathbf{P}_k^+ = \mathbf{P}_{k,N} \tag{13}$$

This last set of equations is clearly wrong because it is equivalent to having N identical but independent measurements. Using these equations the covariance will never converge, but it will keep decreasing because the filter "thinks" more information is available at each iteration.

At least a couple of similar algorithms exist to do line searches in the IKF to insure that the performance index in Eq. (1) decreases at each iteration, alternatively the line search coefficient can be chosen by minimizing Eq. (1) via a nonlinear scalar parameter optimization problem. The iteration is usually terminated when successive iteration produce estimated states that do not differ "much".

THE RECURSIVE UPDATE FILTER

This section introduces the recursive update filter as first derived in [10]. The Kalman filter equations with correlated measurement and process noise [14] are needed. 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 $\mathbf{C}_k$

$$\mathbf{C}_k = \mathrm{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)\boldsymbol{\eta}_k^{\mathrm{T}}\right\}, \tag{14}$$

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

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\eta}_k. \tag{15}$$

Choosing a linear unbiased update

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-). \tag{16}$$

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

$$\mathbf{K}_k = (\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k)(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k\mathbf{C}_k + \mathbf{C}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}})^{-1}, \tag{17}$$

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

$$\begin{aligned}\mathbf{P}_k^+ =& (\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}} - \\ & (\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)\mathbf{C}_k\mathbf{K}_k^{\mathrm{T}} - \mathbf{K}_k\mathbf{C}_k^{\mathrm{T}}(\mathbf{I}_{n\times n} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}},\end{aligned} \tag{18}$$

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

$$\mathbf{e}_k^{(1)} = \mathbf{x}_k - \hat{\mathbf{x}}_k^{(1)} = (\mathbf{I}_{n\times n} - \mathbf{K}_k^{(1)}\mathbf{H}_k)\mathbf{e}_k^- - \mathbf{K}_k^{(1)}\boldsymbol{\eta}_k, \tag{19}$$

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

$$\mathbf{K}_k^{(1)} = \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}. \tag{20}$$

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

$$\mathbf{C}_k^{(1)} = -\mathbf{K}_k\mathbf{R}_k. \tag{21}$$

The updated covariance is obtained simplifying Eq. (18) to obtain

$$\mathbf{P}_k^{(1)} = (\mathbf{I}_{n\times n} - \mathbf{K}_k^{(1)}\mathbf{H}_k)\mathbf{P}_k^- - \mathbf{K}_k\mathbf{C}_k^{(0)\mathrm{T}}. \tag{22}$$

The cross-covariance between the *a priori* state and the measurement error is denoted 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. (22) is only valid when the gain is chosen as the optimal gain, which is not always true in the remaining of this section. Eq. (18) is valid for any choice of $\mathbf{K}_k$.

Processing the same measurement again the second optimal gain is obtained by substituting Eq. (21) into Eq. (17) and replacing $\mathbf{P}_k^-$ with $\mathbf{P}_k^{(1)}$

$$\mathbf{K}_k^{(2)} = (\mathbf{P}_k^{(1)}\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k^{(1)})(\mathbf{H}_k\mathbf{P}_k^{(1)} - \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k\mathbf{C}_k^{(1)} + \mathbf{C}_k^{(1)\mathrm{T}}\mathbf{H}_k^{\mathrm{T}})^{-1}, \tag{23}$$

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

$$\begin{aligned}\mathbf{P}_k^{(1)}\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k^{(1)} &= (\mathbf{I}_{n\times n} - \mathbf{K}_k^{(1)}\mathbf{H}_k)\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \mathbf{K}_k^{(1)}\mathbf{R}_k \\ &= \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \mathbf{K}_k^{(1)}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k) \\ &= \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} - \mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} = \mathbf{O}. \end{aligned} \tag{24}$$

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

$$\mathbf{K}_k^{(1)} = 0.5\,\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}, \tag{25}$$

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.

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

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

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \boldsymbol{\eta}_k \tag{26}$$

the algorithm for the nonlinear recursive update is given by Table 1.

This algorithm is constructed such that, in the linear case, all the steps have the same length, which is reasonable when the number of recursions N is fixed and determined *a priori* . However, notice that there is no requirement for $\gamma_k^{(i)}$ to be chosen as indicated in Table 1. The sequence $\gamma_k^{(i)}$ only has two requirements: it must be composed by number between zero and one (extremes excluded) for all elements except the last one that must be equal to one. Consider the linear case once again and the following coefficients:

$$\gamma_k^{(i)} = \begin{bmatrix} 1/2 & 1/4 & 1 \end{bmatrix} \tag{27}$$

with this choice the first recursion performs half of the update, the second recursion performs one fourth of the other half, or one eighth of the total. The last recursion performs all is left, or three eighths of the total. Since a linear measurement is assumed, these recursions are equivalent to a single update with the Kalman filter.

AUTONOMOUS SELECTION OF ITERATIONS

As mentioned before the IKF is the Gauss-Newton solution to the nonlinear least squares problem given by Eq. (1). In the linear measurement case the performance index becomes:

$$\min_{\mathbf{x}} \mathcal{J} = (\mathbf{x} - \hat{\mathbf{x}}^-)^\mathrm{T}(\mathbf{P}^-)^{-1}(\mathbf{x} - \hat{\mathbf{x}}^-) + (\mathbf{H}\mathbf{x} - \mathbf{y})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}). \tag{28}$$

Similarly to doing a line search in the IKF, an option is to select the number of steps such that each of them guarantees the reduction of the performance index. There are two issues with this approach.

The first is that since the recursive update filter always starts from the previous step, therefore in order to calculate the performance index it would be necessary to invert $\mathbf{P}$ multiple times, which is often a large matrix and usually much larger than the measurement noise or the residual covariance. The second reason is that choosing an updated state guaranteeing that the performance index decreases certainly means a better estimate is found, however it does not guarantee that the state update occurs inside a region where the linearization assumption is valid. The IKF drops the linearization assumption in the state update, however it calculates the covariance still making a first order approximation. The goal of this work is to choose the RUF step size such that each small update occurs in the linear region in order to provide a consistent covariance.

$$\mathcal{J}(\hat{\mathbf{x}}^-) = (\mathbf{H}\hat{\mathbf{x}}^- - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\hat{\mathbf{x}}^- - \mathbf{y}). \tag{29}$$

evaluating it at the posterior state estimate

$$\begin{aligned}
\mathcal{J}(\hat{\mathbf{x}}^+) &= (\mathbf{x}^+ - \hat{\mathbf{x}}^-)^{\mathrm{T}}(\mathbf{P}^-)^{-1}(\mathbf{x}^+ - \hat{\mathbf{x}}^-) + (\mathbf{H}\mathbf{x}^+ - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^+ - \mathbf{y}) && (30)\\
&= \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{P}^-)^{-1}\mathbf{K}\boldsymbol{\epsilon} + (\boldsymbol{\epsilon}^+)^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\epsilon}^+ && (31)
\end{aligned}$$

where $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^-$ is the measurement residual and $\boldsymbol{\epsilon}_k^+$ the post-fit residual. In the linear case $\mathcal{J}(\hat{\mathbf{x}}^-) \geq \mathcal{J}(\hat{\mathbf{x}}^+)$ and that $\boldsymbol{\epsilon}_k^+$ is given by

$$\begin{aligned}
\boldsymbol{\epsilon}_k^+ &= \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^+ = \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^- - \mathbf{H}_k\mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) = (\mathbf{I} - \mathbf{H}_k\mathbf{K}_k)(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) && (32)\\
&= (\mathbf{I} - \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}))(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) && (33)\\
&= \mathbf{R}_k(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) = \mathbf{R}_k\mathbf{W}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) = \mathbf{R}_k\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k^- && (34)
\end{aligned}$$

where $\mathbf{W}_k = \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$, therefore the covariance of $\boldsymbol{\epsilon}_k^+$ is given by

$$\mathbf{W}_k^+ = \mathbf{R}_k\mathbf{W}_k^{-1}\mathbf{R}_k \tag{35}$$

and we have that

$$(\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ = \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k \tag{36}$$

which means that the residual normalized by its standard deviation is the same before and after the update. The posterior performance index is therefore given by

$$\begin{aligned}
\mathcal{J}(\hat{\mathbf{x}}^+) &= \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{P}^-)^{-1}\mathbf{K}\boldsymbol{\epsilon} + (\boldsymbol{\epsilon}^+)^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\epsilon}^+ && (37)\\
&= \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{H}\mathbf{P}^-(\mathbf{P}^-)^{-1}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}\mathbf{W}^{-1}\boldsymbol{\epsilon} + (\boldsymbol{\epsilon}^+)^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{R}\mathbf{R}^{-1}\mathbf{R}\mathbf{W}^{-1}\boldsymbol{\epsilon} && (38)\\
&= \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}\mathbf{W}^{-1}\boldsymbol{\epsilon} + \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{R}\mathbf{W}^{-1}\boldsymbol{\epsilon} && (39)\\
&= \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{W}^{-1}\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ \leq \mathcal{J}(\hat{\mathbf{x}}^-) && (40)
\end{aligned}$$

Notice that the same relation between prior and posterior residual ratios is also valid in the sub-optimal correlated linear case:

$$\begin{aligned}
\boldsymbol{\epsilon}_k^+ &= \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^+ = \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^- - \mathbf{H}_k\mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) = (\mathbf{I} - \mathbf{H}_k\mathbf{K}_k)(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) && (41)\\
&= \Big(\mathbf{I} - \gamma\mathbf{H}_k(\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{C}_k)\,(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k\mathbf{C}_k + \mathbf{C}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}})^{-1}\Big)(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) && (42)\\
&= \Big((1-\gamma)\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \gamma\mathbf{R}_k + (1-\gamma)\mathbf{H}_k\mathbf{C}_k + \gamma\mathbf{C}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\Big)\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k^- = \mathbf{A}_k\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k^- && (43)
\end{aligned}$$

where the residual covariance is now $\mathbf{W}_k = \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k\mathbf{C}_k + \mathbf{C}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}$, therefore the covariance of $\boldsymbol{\epsilon}_k^+$ is given by

$$\mathbf{W}^+ = \mathbf{A}_k\mathbf{W}_k^{-1}\mathbf{A}_k^{\mathrm{T}} \tag{44}$$

and we again have that $(\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ = \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k$.

In the EKF case we have that

$$\boldsymbol{\epsilon}_k^+ = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^+) = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-) - \mathbf{H}_k\mathbf{K}_k(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-)) + \boldsymbol{\delta} = \mathbf{A}_k(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-)) + \boldsymbol{\delta} \tag{45}$$

where $\boldsymbol{\delta}$ is a vector containing the higher order terms of the Taylor series expansion of the measurement function. We then have that

$$(\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ = \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k + \kappa \tag{46}$$

where κ is some scalar function of the higher order terms. In the linear case we know in advance what the value for the updated normalized residual is, while in the nonlinear case $(\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ \simeq \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k$ only when the higher order effects of the measurement nonlinearity are negligible. We can take advantage of this knowledge to select the length of of the update step. Furthermore by insuring that $(\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ \simeq \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k$ we can also guarantee that the updated state has a smaller performance index. The algorithm to adaptably determine the number of recursions N is summarized in Table 2

A few comments are in order. In Table 2 the positive scalar ϑ is a design variable indicating the amount of nonlinear effects tolerated while selected N. A value of zero means $(\boldsymbol{\epsilon}_k^+)^{\mathrm{T}}(\mathbf{W}_k^+)^{-1}\boldsymbol{\epsilon}_k^+ = \boldsymbol{\epsilon}_k^{\mathrm{T}}\mathbf{W}_k^{-1}\boldsymbol{\epsilon}_k$ is strictly enforced. Selecting $\vartheta = 0$ is a poor choice for nonlinear systems as the algorithm will never reach its exit condition. A ceiling to the maximum number of iterations is giving by N_{max}, it is a standard practice to assure an algorithm will not loop forever, although this condition should never be exercised with an appropriate value of ϑ. This algorithm selects the number of recursions based on the length of the very first recursion. All of the successive recursions check that the current length is appropriate, if not the length is decreased. Assume for example that during the first recursion $N = 2$ is selected. Therefore right before entering the second recursion we set $N_{min} = N = 2$. The second recursion ($i = 2$) starts with $N = 2$ and calculates $\gamma_k^{(2)} = 1$, however assume the check fails and therefore overwrites N as $N = 3$, which makes $\gamma_k^{(2)} = 0.5$. A third recursion is therefore called with $\gamma_k^{(3)} = 1$.

Notice that $\mathbf{W}_k^{(i)}$ is assumed invertible. This condition is always satisfied so long as $\mathbf{R}$ is invertible.

NUMERICAL RESULTS

The example presented here is spacecraft rendezvous as shown in Fig. 1. The chaser spacecraft starts the approach one hundred meters in front of the target performing a +V-bar approach. 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. The dynamics is governed by the CW equations

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \boldsymbol{\nu} \qquad\qquad \mathbf{P}_0 = \begin{bmatrix} 10^2\mathbf{I}_{3\times 3} & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{3\times 3} & 0.05^2\mathbf{I}_{3\times 3} \end{bmatrix} \tag{47}$$

Table 2. Autonomous selection of N, Recursive Update Filter

```
$N = 1, N_{min} = 1, i = 1$
$\mathbf{C}_k^{(0)} = \mathbf{O},\ \mathbf{P}_k^{(0)} = \mathbf{P}_k^-,\ \hat{\mathbf{x}}_k^{(0)} = \hat{\mathbf{x}}_k^-$
while $i = 1 \leq N$
  $\mathbf{H}_k^{(i)} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{x}}_k^{(i-1)}}$
  $\mathbf{W}_k^{(i)} = \mathbf{H}_k^{(i)}\mathbf{P}_k^{(i-1)}\mathbf{H}_k^{(i)\mathrm{T}} + \mathbf{R}_k + \mathbf{H}_k^{(i)}\mathbf{C}_k^{(i-1)} + \mathbf{C}_k^{(i-1)\mathrm{T}}\mathbf{H}_k^{(i)\mathrm{T}}$
  for $N = N_{min}$ to $N_{max}$
    $\gamma_k^{(i)} = 1/(N+1-i)$
    $\boldsymbol{\epsilon}_k^{(i-1)} = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^{(i-1)})$
    $\mathbf{K}_k^{(i)} = \gamma_k\big(\mathbf{P}_k^{(i-1)}\mathbf{H}_k^{(i)\mathrm{T}} + \mathbf{C}_k^{(i-1)}\big)\big(\mathbf{W}_k^{(i)}\big)^{-1}$
    $\hat{\mathbf{x}}_k^{(i)} = \hat{\mathbf{x}}_k^{(i-1)} + \mathbf{K}_k^{(i)}\boldsymbol{\epsilon}_k^{(i)}$
    $\boldsymbol{\epsilon}_k^{(i)} = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^{(i)})$
    $\mathbf{A}_k^{(i)} = (1-\gamma)\mathbf{H}_k^{(i)}\mathbf{P}_k^{(i-1)}\mathbf{H}_k^{(i)\mathrm{T}} + \gamma\mathbf{R}_k + (1-\gamma)\mathbf{H}_k^{(i)}\mathbf{C}_k^{(i-1)} + \gamma\mathbf{C}_k^{(i-1)\mathrm{T}}\mathbf{H}_k^{(i)\mathrm{T}}$
    $\mathbf{W}_k^{(i)} = \mathbf{A}_k^{(i)}(\mathbf{W}_k^{(i-1)})^{-1}\mathbf{A}_k^{(i)\mathrm{T}}$
    if $\Big|(\boldsymbol{\epsilon}_k^{(i)})^{\mathrm{T}}(\mathbf{W}_k^{(i)})^{-1}\boldsymbol{\epsilon}_k^{(i)} - (\boldsymbol{\epsilon}_k^{(i-1)})^{\mathrm{T}}(\mathbf{W}_k^{(i-1)})^{-1}\boldsymbol{\epsilon}_k^{(i-1)}\Big| / [(\boldsymbol{\epsilon}_k^{(i-1)})^{\mathrm{T}}(\mathbf{W}_k^{(i-1)})^{-1}\boldsymbol{\epsilon}_k^{(i-1)}] \leq \vartheta$
    then
      EXIT THE FOR LOOP
    end if
  end for
  $\mathbf{K}_k^{(i)} = \gamma_k^{(i)}\big(\mathbf{P}_k^{(i-1)}\mathbf{H}_k^{(i)\mathrm{T}} + \mathbf{C}_k^{(i-1)}\big)\big(\mathbf{W}_k^{(i)}\big)^{-1}$
  $\hat{\mathbf{x}}_k^{(i)} = \hat{\mathbf{x}}_k^{(i-1)} + \mathbf{K}_k^{(i)}\big(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_k^{(i-1)})\big)$
  $\mathbf{P}_k^{(i)} = \big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)\mathbf{P}_k^{(i-1)}\big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)^{\mathrm{T}} + \mathbf{K}_k^{(i)}\mathbf{R}_k\mathbf{K}_k^{(i)\mathrm{T}} - \big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)\mathbf{C}_k^{(i-1)}\mathbf{K}_k^{(i)\mathrm{T}} - \mathbf{K}_k^{(i)}\mathbf{C}_k^{(i-1)\mathrm{T}}\big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)^{\mathrm{T}}$
  $\mathbf{C}_k^{(i)} = \big(\mathbf{I}_{n\times n} - \mathbf{K}_k^{(i)}\mathbf{H}_k^{(i)}\big)\mathbf{C}_k^{(i-1)} - \mathbf{K}_k^{(i)}\mathbf{R}_k$
  $i = i+1, N_{min} = N$
end while
$\mathbf{P}_k^+ = \mathbf{P}_k^{(N)},\ \hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^{(N)}$
```

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{O}_{3\times3} & \mathbf{I}_{3\times3} \\ \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \tag{48}$$

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 is the target's orbital rate. The process noise is given by

$$\mathrm{E}\{\boldsymbol{\nu}\} = \mathbf{0} \qquad \mathrm{E}\{\boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{T}}\} = \begin{bmatrix} \mathbf{O}_{3\times3} & \mathbf{O}_{3\times3} \\ \mathbf{O}_{3\times3} & 10^{-9}\mathbf{I}_{3\times3} \end{bmatrix} \tag{49}$$

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

$\epsilon = \sin^{-1}(\mathbf{x}(3)/\rho)$.

$$\mathbf{y} = \begin{bmatrix} \rho \\ \alpha \\ \epsilon \end{bmatrix} + \boldsymbol{\eta} \tag{50}$$

$$\mathrm{E}\left\{\boldsymbol{\eta}\right\} = \mathbf{0}, \qquad \mathrm{E}\left\{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}\right\} = \begin{bmatrix} 0.1^2 & \mathbf{O}_{1\times 2} \\ \mathbf{O}_{2\times 1} & (0.1\pi/180)^2 \mathbf{I}_{2\times 2} \end{bmatrix} \tag{51}$$

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. 2 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 [9].

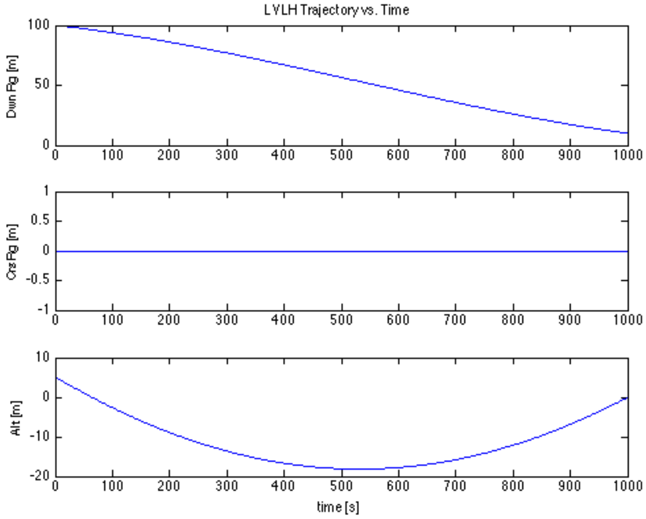


Figure 1. Trajectory.

Fig. 3 shows the performance of 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. 2 and 3. The algorithm selects the number of recursions to be equal to one for at all times except the very first one. The number of recursions at the first step varies between 3 and 10 depending on the particular run, as seen in Fig. 4

CONCLUSIONS

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

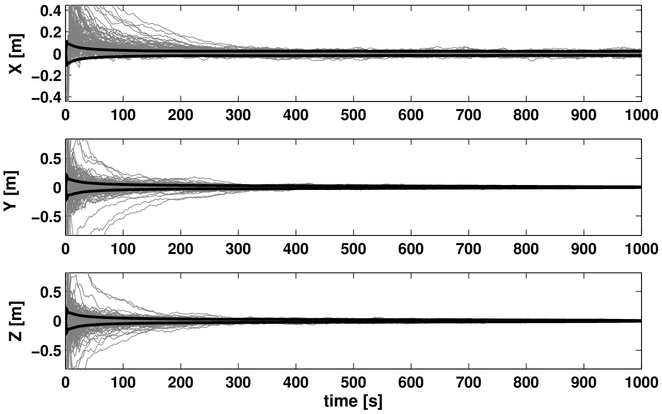


Figure 2. EKF Estimation Error and Predicted Covariance.

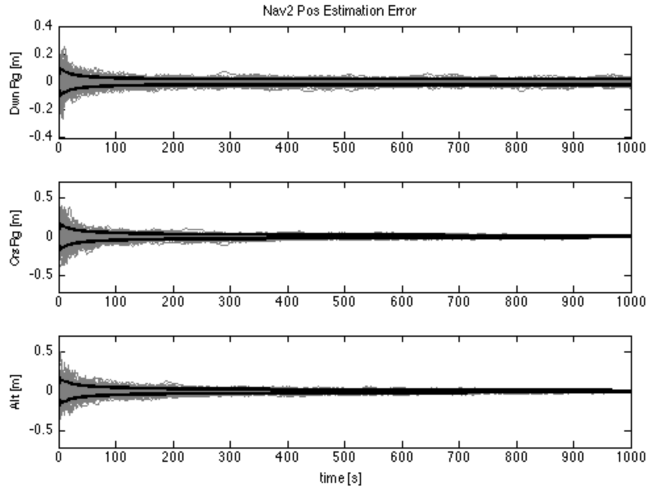


Figure 3. Recursive Update Filter Estimation Error and Predicted Covariance.

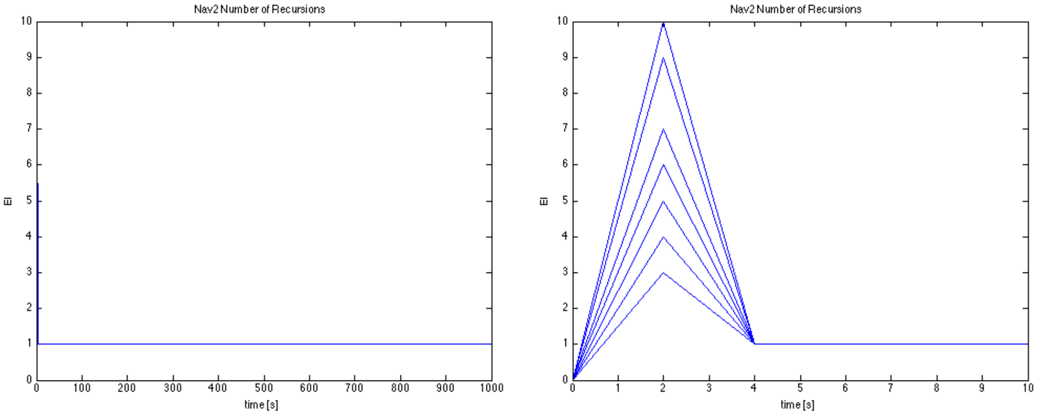
(i.e. the estimate without incorporating the measurements) but there is no guarantee it performs better than the measurements themselves. Nonlinear estimators such as the iterated Kalman filter or the recursive update filter outperform the extended Kalman filter (EKF) in the presence of nonlinear measurements. This paper introduces a novel algorithm to adaptably and autonomously select the number of recursions on the go. The algorithm takes advantage of the relation between the measurement residuals and their covariance to keep each recursion in an area where the linearization assumption holds.

REFERENCES

[1] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, March 1960, doi:10.1115/1.3662552.

[2] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction,” *Journal of Basic Engineering*, vol. 83, no. Series D, pp. 95–108, March 1961, doi:10.1115/1.3658902.

[3] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: The MIT press, 1974.



(a) Zoom-Out View

(b) Zoom-In View

Figure 4. Number of Recursions Selected by Algorithm.

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

[5] 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, doi: 10.1109/SAP.1970.270017.

[6] 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.

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

[8] 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, doi: 10.1109/9.847726.

[9] 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.

[10] R. Zanetti, “Recursive Update Filtering for Nonlinear Estimation,” *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1481–1490, June 2012, doi:10.1109/TAC.2011.2178334.

[11] 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, 10.1109/9.250476 .

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

[13] X.-l. Niu, G.-q. Zhao, Y.-h. Liu, and H. Chang, “An improvement on the iterated kalman filter,” in *Radar Conference, 2009 IET International*, 2009, pp. 1–4.

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