Discrete-time Unscented Kalman Filter: Comprehensive Study of Stochastic Stability*

Gyorgyi Dymirkovsky, Juanwei Ying, and Jiahe Xu

Abstract—The performance of the Unscented Kalman Filter (UKF) for a class of general nonlinear stochastic discrete-time systems is investigated in this paper. It is proved that the estimation error of the UKF remains bounded provided a under certain conditions are satisfied. It is further shown that the estimation error remains bounded provided the system satisfies the nonlinear observability rank condition. Furthermore, it is shown that the design of noise covariance matrix plays an important role in improving the stability of the UKF algorithm. These results are verified by simulations for a given illustrative example of an inherently nonlinear plant.

I. INTRODUCTION

State estimation and Kalman filtering never seized to attract considerable research efforts worldwide since the famous seminal papers [1], [2] by R.E. Kalman based on new insights into dynamic systems and following his new approach to general theory of control systems [2], [3]. These new insights involved not only concept of systems state but also system mechanism properties such as observability and controllability in addition to mechanisms of stability and instability. A great many researchers have extended these ideas and insights, and some of recent ones are found in [5-8]. In a rather condensed summary these may be envisaged as in Figure 1.

It has to be immediately emphasised, however, Kalman filtering has transcended into a unique discipline in systems and signal processing sciences, which yielded a number of successful technologies for various applications [9]. Although initially introduced and developed for estimation problems in linear systems, it has been extended in due time to nonlinear systems within the context of Extended Kalman Filters (EKF) [9-14]. Moreover, articles by R. Unbehauen and co-authors [15-17] provided the first analytical results on stochastic stability for both discrete-time and continuous-time EKF. This way they contributed sound foundations for the EKF system theory for nonlinear estimation applications. Nonetheless, the issues of state estimation and Kalman filtering for nonlinear systems remained open to further research because of the phenomena uniqueness in nonlinear dynamic processes.

In early 1990-ties Julier and co-authors proposed the Unscented Kalman Filter (UKF) in [18], [20] precisely for the essential extension of Kalman filtering to nonlinear estimation problems. They also showed the UKF was a considerable improvement in comparison with the EKF [18], [20], [21], [24]. The UKF is based on employing the special transformation technique that is a mechanism for propagating mean and covariance through a nonlinear transformation [19], [25] and called unscented transformation (UT). Thus it is no longer necessary to use a linearization technique and compute the system Jacobian and Hessian matrices for the UKF [20-24]. This has enabled essential avoidance of the error produced by the interruption of higher-order terms and the precision can reach the second-order even higher, e.g. as precise as third-order to the Gauss noise [20]. By nowadays, the UKF is widely used in various applications, ranging from target tracking [18] to position determination [23], multi-sensor fusion [26], estimation in flight control under wind shear [27], etc. The inventor Julier has compared the performances of both the UKF and the EKF for an example nonlinear system, as in this paper, and he showed that the UKF outperforms considerably the EKF [21]. Similarly, a comparison has been established in [26]. These findings are further supported with the published results in [28-32] as well as [36].

![Figure 1. Mapping structure of dynamic system processors in engineering terms and consistency with the semi-group algebraic setting [3, 5, 41].](image)

It should be noted, nonetheless, superior performance of the UKF and its practical usefulness is accompanied with a certain heuristics in its original theoretical derivation [18-20], which makes rather difficult mathematically rigorous derivations of unscented Kalman filters for various applications. Furthermore, the properties of stability and convergence for the UKF are considerably hard to analyze hence have been developed only for special applications.
where the considered nonlinear systems are assumed along with linear measurement equation [28].

In addition to these results, one of the two recent research directions is towards studying extensions of the UKF to operation circumstances with intermittent observations [34], [35], [37] or in the presence of packet dropouts for the case of discrete-time systems [36]. The other direction is in studying the UKF for a more general nonlinear case in a stochastic framework [11], [38], which remains of primary interest given the variety of potential applications. Moreover, some interesting relationships between the observability of nonlinear systems [39], [40] and the detectability of time-varying linear systems [41], [34] in conjunction with the existence of positive definite solutions for the UKF have also emerged as an important research area.

The here reported research was motivated by the encouraging results on stability analysis both for the standard and extended Kalman filtering [9], [13] as well as on stochastic stability analysis for more general nonlinear estimation problems [12], [15-17]. Additional motivation was the recently solved continuous-time UKF case in [33]. Thus, via studying the dynamics of the discrete-time UKF, this paper presents the relevant result on stochastic stability and its underlying relation to nonlinear system observability are derived. The main contribution of this paper is the proof that, under certain conditions, the estimation error of the UKF remains bounded in the sense of mean square. In order to improve the stability, slight modifications of the standard UKF were performed by introducing an additive, positive definite matrix into the noise covariance matrix. It is shown that if this extra matrix is properly selected then the performance of the UKF used for general nonlinear systems may be improved significantly even in the presence of big initial estimation error. Furthermore, the role of nonlinear observability in this context is also established. The modified UKF is applied as an estimator to a relevant example from the literature in order to illustrate the applicability of the new theoretical results as well as to seek verification via computer simulations.

Further the paper is written as follows. In Section II, and overview of the UKF is given. In Section III, the stability analysis is carried out and the first new result presented along with its proof. In Section IV, the second new result is given along with its proof which points out the importance of nonlinear systems observability for the UKF design. Section V presents the illustrative example and discusses the obtained simulation results in the necessary detail. Concluding remarks and references are given thereafter.

II. THE DISCRETE-TIME UNSCENTED KALMAN FILTER

The study technique employed in this paper was inspired by the work of Reif and co-authors in [16] for discrete-time extended Kalman filter. In addition, the considered class of fairly general nonlinear discrete-time systems is assumed to be represented by

\[
\begin{align*}
    x_k &= f(x_{k-1}) + G_k w_{k-1}, \\
    y_k &= h(x_k) + D_k v_k
\end{align*}
\]

where, \( k \in N \) discrete time, \( N \) denotes the set of natural numbers including zero. In (1), \( x_k \in R^r \) represents the state and \( y_k \in R^m \) the measurement. Nonlinear functions \( f(\cdot) \) and \( h(\cdot) \) are assumed to be continuously differentiable with respect to \( x_k, w_k \) and \( v_k \). The latter two represent \( R^k \) and \( R^l \) vector-valued, uncorrelated, zero-mean white noise processes with identity covariance. It is assumed \( x_0 \) is uncorrelated with \( w_k \) and \( v_k \), and \( E(x_0) = \tilde{x}_0 \), \( \text{cov}(x_0) = P_0 \). The variances of \( w_k \) and \( v_k \) satisfy the following expressions

\[
E[w_i w_j^T] = \begin{cases} Q_k & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}, \quad E[v_i v_j^T] = \begin{cases} R_k & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}
\]

where, \( Q_k \) is the system’s noise sequence covariance matrix, and it is a symmetrical non-negative definite matrix. Matrix \( R_k \) is measurement noise sequence covariance matrix and it is a symmetrical positive definite matrix.

The procedure for implementing the UKF, on the grounds of source [25], can be summarized as follows.

- The n-dimensional random variable \( x_k \) with mean \( \tilde{x}_k \) and covariance \( \hat{P}_k \) can be approximated by sigma points \( x_{i,k} \) selected from the columns of \( \tilde{x}_{k-1} \pm a \sqrt{n \hat{P}_{k-1}} \), \( i = 0, \ldots, 2n \). The opposite weights are \( \omega_0 = 1 - \frac{1}{2a^2} \), \( \omega_i = 1/2n a^2 \), \( i = 1, 2, \ldots, 2n \).

- Each point is instantiated through the process model to yield a set of transformed samples; the predicted mean and covariance are computed as

\[
\begin{align*}
    x_{i,k+1} &= f(x_{i,k}) + G_k x_{i,k} + \xi_{i,k} \\
    P_{i,k+1} &= \mathbf{G}_k P_{i,k} \mathbf{G}_k^T + \xi_{i,k} \xi_{i,k}^T + \Delta Q_k
\end{align*}
\]

where, \( \Delta Q_k \) is an extra positive definite matrix introduced in the calculated covariance matrix as a slight modification of the UKF so that the stability will be improved.

- Then the measurement update can be performed with the equations as follows.

\[
\begin{align*}
    y_{i,k+1} &= h(x_{i,k+1}) \\
    \hat{P}_{i,y} &= \sum_{i=0}^{2n} \omega_i (y_{i,k+1} - \hat{y}_k)(y_{i,k+1} - \hat{y}_k)^T + D_k R_k D_k^T \\
    \hat{P}_y &= \sum_{i=0}^{2n} \omega_i (\tilde{y}_{i,k+1} - \hat{y}_k)(\tilde{y}_{i,k+1} - \hat{y}_k)^T \\
    W_k &= \hat{P}_y^{-1} \hat{P}_{i,y}^{-1}
\end{align*}
\]
\[ \hat{x}_k = \delta_{k-1} + W_x (\hat{y}_k - \hat{y}_k) \tag{8} \]
\[ \hat{P}_k = \hat{P}_{k-1} + W_{k} \hat{P}_{k-1} W_{k}^T \tag{9} \]

Clearly, the implementation of the UKF appears extremely convenient because there is no need to evaluate the Jacobian matrices, which is necessary in the case of the EKF.

III. STABILITY ANALYSIS OF THE UKF

In this section, a simple approach to represent the error dynamics of the UKF for general nonlinear systems is given and its stochastic boundedness is established and proved.

A. Instrumental Diagonal Matrix and an Extra Positive Definite Matrix

First instrumental time-varying matrices are introduced in order to give a formulation for the UT technique of Julier and his co-authors. Define the estimation error and prediction order to give a formulation for the UT technique of Julier and its stochastic boundedness is established and proved.

\[ \bar{x}_k = x_k - \hat{x}_k \tag{10} \]
\[ \bar{x}_{k+1|k} = x_{k+1} - \hat{x}_{k+1|k} \tag{11} \]

Expanding \( x_k \) in (1) by means of a Taylor series about \( x_k \) gives,

\[ x_k = f(\hat{x}_{k-1}) + \nabla f(\hat{x}_{k-1}) \delta_{k-1} + \frac{1}{2} \nabla^2 f(\hat{x}_{k-1}) \delta_{k-1}^2 + \ldots + G_{w_k} \delta_{k-1} \tag{12} \]

where \( \nabla^2 f(\hat{x}_{k-1}) = \left( \frac{\partial^2 f}{\partial x_j \partial x_j} \right) \), \( x_j \) denotes the \( j \)-th component of \( x \). Expanding \( \hat{x}_{k|k} \) given in (2) by a Taylor series yields,

\[ \hat{x}_{k|k} = \left( 1 - \frac{1}{a} \right) f(\hat{x}_{k-1}) + \frac{1}{2a} \sum_{j=1}^{k} \left[ \hat{x}_{k-1} + \left( a \sqrt{L_{k-1}} \hat{P}_{k-1} \right) \right] \nabla f(\hat{x}_{k-1}) x_j \]
\[ + \frac{1}{2a} \sum_{j=1}^{k} \left[ \hat{x}_{k-1} + \left( a \sqrt{L_{k-1}} \hat{P}_{k-1} \right) \right] \nabla^2 f(\hat{x}_{k-1}) x_j^2 \tag{13} \]

\[ = f(\hat{x}_{k-1}) + \frac{1}{2} \nabla^2 f(\hat{x}_{k-1}) \delta_{k-1}^2 + \ldots \]

The substitution of (12) and (13) into (11) gives an approximate equality

\[ \bar{x}_{k|k} \approx G_{k} \delta_{k-1} + \hat{P}_{k} \delta_{k-1} + \frac{\partial f(x)}{\partial x} \delta_{k-1} \tag{14} \]

In (14), it is evident that there always exist residuals of state error prediction \( \bar{x}_{k|k} \). In order to take these residuals into account and obtain a more exact equality, an unknown instrumental diagonal matrix \( A_{k} = \text{diag}(\lambda_{k}, \lambda_{k}, \ldots, \lambda_{M,k}) \) is introduced, so that

\[ \bar{x}_{k|k} = \lambda_{k} \hat{P}_{k} x_{k} + G_{k} \delta_{k-1} \tag{15} \]

The residual of the measurement can also be defined by

\[ \bar{y}_k = y_k - \hat{y}_k = \Gamma_k H_k \bar{x}_{k|k} + D_k y_k \tag{16} \]
B. Stochastic Boundedness of the Estimation Error Dynamics

For the analysis of the error dynamics some standard results about the boundedness of stochastic processes from [11], [38] are recalled.

Lemma 3.1: Assume that $\zeta_k$ is the stochastic process and there is a stochastic process $V(\zeta_k)$ as well as real numbers $\nu_{\min}, \nu_{\max} > 0$, $\mu, \alpha > 0$, and $0 < \alpha \leq 1$ such that for any $k$

$$
\nu_{\min} \| \zeta_k \| \leq V(\zeta_k) \leq \nu_{\max} \| \zeta_k \|
$$

are fulfilled. Then the stochastic process is bounded in the mean square, that is

$$
E\| \zeta_k \|^2 < \nu_{\max} E \| \zeta_0 \|^2 + \mu \sum_{j=1}^{k} \| 1 - \alpha \|^j.
$$

(24)

For the purpose of establishing the sufficient conditions that ensure stability of the UKF also another two lemmas, given below, are needed.

Lemma 3.2: Assume that matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times m}$, if $A > 0$ and $C > 0$, then

$$
A^{-1} > (B^T AB + C)^{-1} B^T
$$

(27)

Lemma 3.3: Assume that matrices $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{p \times n}$, if $A > 0$ and $C > 0$, then

$$
A^{-1} > (A + C)^{-1}
$$

(28)

With Lemmas 3.1 – 3.3 and the formulations shown in (15), (16), (18) and (21), it becomes possible to state the first main result of this paper.

Theorem 3.1: Consider general nonlinear stochastic systems as represented by (1) and the UKF described by (2)-(9). Further, suppose the following assumptions hold true:

1. There are real numbers $f_{\min}, h_{\min}, \lambda_{\min}, \gamma_{\min} \neq 0$, and $f_{\max}, h_{\max}, \lambda_{\max}, \gamma_{\max} \neq 0$, such that the following bounds on various matrices are fulfilled for every $k \geq 0$:

$$
\begin{align*}
\nu_{\min} f_k &\leq f_k f_k^T \nu_{\max}, \\
\nu_{\min} h_k &\leq h_k h_k^T \nu_{\max}, \\
\nu_{\min} \lambda_k &\leq \lambda_k \lambda_k^T \nu_{\max}, \\
\nu_{\min} \gamma_k &\leq \gamma_k \gamma_k^T \nu_{\max}.
\end{align*}
$$

(29)

(30)

(31)

(32)

2. There are real numbers $q_{\min}, q_{\max}, \hat{q}_{\min}, \hat{q}_{\max}, \hat{r}_{\min}, \hat{r}_{\max}, p_{\min}, p_{\max}, \hat{p}_{\min} > 0$, such that the following bounds are fulfilled:

$$
\begin{align*}
p_{\min} l &\leq \hat{p}_k \leq p_{\max} l, \\
q_{\min} l &\leq Q_k \leq q_{\max} l, \\
\hat{q}_{\min} l &\leq \hat{Q}_k \leq \hat{q}_{\max} l.
\end{align*}
$$

(33)

(34)

(35)

Then the estimation error $\tilde{x}_k$ given by (10) is exponentially bounded in the mean square.

Proof: Due to Lemma 3.1, the following Lyapunov quadratic form is chosen

$$
V_k(\tilde{x}_k) = \tilde{x}_k^T \tilde{P}_k^{-1} \tilde{x}_k
$$

(37)

From (33) it follows

$$
\frac{1}{p_{\max}} V_k(\tilde{x}_k) \leq V(\tilde{x}_k) \leq \frac{1}{p_{\min}} V_k(\tilde{x}_k)
$$

(38)

In order to satisfy the requirement for Lemma 3.1 to be applied, it is necessary to have an upper bound on $E[V(\zeta_k)] - V(\zeta_k)$. Thus substitution of (21) and (23) into (9) yields

$$
\tilde{r}_k = \tilde{p}_{k,1} - \tilde{p}_{k,2} \tilde{x}_k \tilde{p}_{k,1}^T = (I - W_k \Gamma_k) \tilde{r}_{k,1}
$$

(39)

where

$$
W_k = \tilde{p}_{k,1} \Gamma_k \tilde{p}_{k,1}^T + \tilde{r}_{k,1} - \tilde{x}_k \tilde{x}_k^T
$$

By making use of (8), (10), (16) as well as (40) it can be shown

$$
\tilde{x}_k = \tilde{x}_k - \tilde{r}_{k,1} - \tilde{r}_{k,2} - \tilde{x}_k \tilde{x}_k^T
$$

(41)

Then from (37) and (41) it can be found

$$
\begin{align*}
V_k(\tilde{x}_k) &= \tilde{x}_k^T \tilde{x}_k - \tilde{x}_k \tilde{x}_k^T \tilde{r}_{k,1} - \tilde{x}_k \tilde{x}_k^T \tilde{r}_{k,2} - \tilde{x}_k \tilde{x}_k^T \\
&= \tilde{x}_k^T \tilde{x}_k - \tilde{x}_k \tilde{x}_k^T \tilde{r}_{k,1} - \tilde{x}_k \tilde{x}_k^T \tilde{r}_{k,2} - \tilde{x}_k \tilde{x}_k^T \\
&= \tilde{x}_k^T \tilde{x}_k - \tilde{x}_k \tilde{x}_k^T \tilde{r}_{k,1} - \tilde{x}_k \tilde{x}_k^T \tilde{r}_{k,2} - \tilde{x}_k \tilde{x}_k^T.
\end{align*}
$$

(42)

Rearranging (40) yields

$$
W_k = (I - \Gamma_k H_k) \tilde{p}_{k,1} \Gamma_k \tilde{r}_{k,1} - \tilde{p}_{k,1} \Gamma_k \tilde{r}_{k,1} - \tilde{p}_{k,1} \Gamma_k \tilde{r}_{k,1} - \tilde{p}_{k,1} \Gamma_k \tilde{r}_{k,1}
$$

(43)

On the other hand, applying the well known matrix inversion lemma on (39) gives

$$
\tilde{P}_k^{-1} = \tilde{P}_{k,1} \Gamma_k \tilde{r}_{k,1} H_k
$$

(44)

Via inserting these (43), (44) and also (15) into (42), and then taking the conditional expectation, it follows:

$$
E[V(\zeta_k)|x(k-1)]=
$$

(45)

Next, the term $(\tilde{r}_k - \tilde{r}_k)^T \Gamma_k H_k \tilde{P}_k H_k^T \tilde{r}_k$ on the right side of (45) is to closely examine. By using (43) and (40) it can be verified...
(\hat{R}_k - \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \hat{R}_k^{-1})
\quad = \hat{R}_k^{-1} [1 - \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k] (\Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \hat{R}_k^{-1} - 1)] (46)

By substituting (46) and (18) into (45), and then applying Lemma 3.3, the expectation (45) becomes:

\[ E[V_k(x_{k-1})] \leq E[V_k(x_{k-1})] = E\left[ \left( \alpha \gamma_{max} + \gamma_{min} \right) \left( \alpha \gamma_{max} + \gamma_{min} \right) \right] \]

Notice now the inequalities (29) and (31) imply that \((\Lambda_k F_k)^{-1}\) exists. It is therefore that it may be established

\[ E\left[ \left( \alpha \gamma_{max} + \gamma_{min} \right) \left( \alpha \gamma_{max} + \gamma_{min} \right) \right] = \gamma_{min} \gamma_{max} = \gamma_{max} \gamma_{max} \]

Subtraction of (48) from both sides of (47) then yields

\[ E[V_k(x_{k-1})] \leq E[V_k(x_{k-1})] = E\left[ \left( \alpha \gamma_{max} + \gamma_{min} \right) \left( \alpha \gamma_{max} + \gamma_{min} \right) \right] \]

Now, it is the last term in (49) to be closely examined. According to Lemma 3.2 it may well be inferred

\[ \hat{P}_{k-1} = \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) ^{-1} \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \hat{P}_k H_k^\dagger \Gamma_k \right) ^{-1} \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) \]

By means of pre- and post-multiplication of both sides of (50) by \(x_{k-1}\) and \(x_{k-1}\), respectively, for \(\forall x_{k-1} \neq 0\) it holds

\[ x_{k-1}^\dagger \hat{P}_{k-1} x_{k-1} > x_{k-1}^\dagger \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) x_{k-1} \]

From (51), upon denoting

\[ \alpha_k = \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) \]

it follows that \(\alpha_k < 1\).

Under the assumed inequalities (29)-(36), on the other hand, it follows

\[ \alpha_k \geq p_{max} \left( \gamma_{min} \alpha_{min} \gamma_{min} f_{min} \right) \left( p_{max} \left( \gamma_{max} \gamma_{max} \gamma_{max} f_{max} \right) + \gamma_{max} \gamma_{max} f_{max} \right) \]

Thus, by using (51) and (53) it is readily shown that

\[ \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) \left( \Gamma_k H_k \hat{P}_k H_k^\dagger \Gamma_k \right) \]
Remark 2. To ensure the stability of the UKF, matrices \( \hat{Q}_k \) need to be positive definite. On the grounds of (19), as \( AP_{k:k-1} \) and \( \partial \Phi_{k:k-1} \) may be not positive definite matrices, an extra additive matrix \( \Delta Q_k \) should be introduced to modify the UKF slightly so that \( \hat{Q}_k \equiv \hat{Q}_k \) be satisfied always. Obviously, if \( \Delta Q_k \) is sufficiently large, condition (35) can always be fulfilled. This means that the UKF can tolerate high order error introduced during the UT by enlarging the noise covariance matrix. On the other hand, the precision of the algorithm also is related to the value of \( \hat{Q}_k \).

Remark 3. Condition (33) is closely related to the observability property of the general nonlinear system (1) as the related discussion in the next Section IV shows.

IV. NONLINEAR SYSTEMS OBSERVABILITY IS SIGNIFICANT FOR THE UKF DESIGNS

In this section, the related observability property of the general nonlinear system is more closely discussed in conjunction with the stochastic stability of the UKF. For this purpose, firstly the following results on the observability rank condition for every vector \( x_k \in \mathbb{R}^r \), and an integer \( k \geq 0 \) and let the observability Gramian be given by

\[
U(x_k) = \begin{bmatrix}
\frac{\partial h}{\partial x}(x_k) \\
\frac{\partial h}{\partial x}(x_{k+1}) \frac{\partial f}{\partial x}(x_k) \\
\vdots \\
\frac{\partial h}{\partial x}(x_{k+r-1}) \frac{\partial f}{\partial x}(x_{k-r+2}) \frac{\partial f}{\partial x}(x_k)
\end{bmatrix}
\]  

(60)

has full rank \( r \) at \( x_k \).

For the proof of this theorem we make use of some auxiliary results. First we recall the uniform observability of linear time-varying systems [41].

Lemma 4.1: The general nonlinear system given by (1) satisfies the nonlinear observability rank condition at \( x_k \in \mathbb{R}^r \), if the nonlinear observability matrix

\[
U(x_k) = \begin{bmatrix}
\frac{\partial h}{\partial x}(x_k) \\
\frac{\partial h}{\partial x}(x_{k+1}) \frac{\partial f}{\partial x}(x_k) \\
\vdots \\
\frac{\partial h}{\partial x}(x_{k+r-1}) \frac{\partial f}{\partial x}(x_{k-r+2}) \frac{\partial f}{\partial x}(x_k)
\end{bmatrix}
\]  

(60)

has full rank \( r \) at \( x_k \).

Lemma 4.2: Consider time-varying matrices \( \Lambda_k F_i, \Gamma_i H_k, \forall k \geq 0 \) and let the observability Gramian be given by

\[
M_{k+1:k} = \sum_{i=k}^{k+r-1} \Phi_i \Gamma_i H_i \Phi_{i:k}
\]  

(61)

for some integer \( n > 0 \) with \( \Phi_{i:k} = I \) and

\[
\Phi_{i:k} = \Lambda_{i-1} F_{i-1} \cdots I
\]  

(62)

for \( i > k \). Then matrices \( \Lambda_k F_i, \Gamma_i H_k, \forall k \geq 0 \) are said to satisfy the uniform observability condition, if there are real numbers \( m_{\min}, m_{\max} > 0 \) and an integer \( l > 0 \), such that the following inequality holds:

\[
m_{\max} l \leq M_{k+1:k} \leq m_{\min} l
\]  

(63)

Lemma 4.3: Consider the measurement covariance \( \hat{P}_k \) for \( \forall k \geq 0 \), and let the following conditions hold:

(1) There are real numbers \( q_{\min}, q_{\max}, \hat{q}_{\min}, \hat{q}_{\max}, \hat{r}_{\min}, \hat{r}_{\max} > 0 \) such that the matrices \( Q_k, \hat{Q}_k \) and \( \hat{R}_k \) are bounded by

\[
q_{\min} l \leq Q_k \leq q_{\max} l
\]  

(64)

\[
\hat{q}_{\min} l \leq \hat{Q}_k \leq \hat{q}_{\max} l
\]  

(65)

\[
\hat{r}_{\min} l \leq \hat{R}_k \leq \hat{r}_{\max} l
\]  

(66)

(2) Matrices \( \Lambda_k F_k, \Gamma_k H_k \) satisfy the uniform observability condition.

(3) The initial condition matrix \( P_0 \) is positive definite.

Then there exist positive numbers \( p_{\max}, p_{\min} > 0 \) such that \( \hat{P}_k \) is bounded via

\[
p_{\min} l \leq \hat{P}_k \leq p_{\max} l
\]  

(67)

for every \( k \geq 0 \).

Now it is possible to state the other main new result.

Theorem 4.1: Consider the general nonlinear stochastic systems represented by (1) and the UKF as described by (2)-(9). Assume there exist real numbers \( q_{\min}, q_{\max}, \hat{q}_{\min}, \hat{q}_{\max}, \hat{r}_{\min}, \hat{r}_{\max} > 0 \) satisfying

\[
q_{\min} l \leq Q_k \leq q_{\max} l
\]  

(68)

\[
\hat{q}_{\min} l \leq \hat{Q}_k \leq \hat{q}_{\max} l
\]  

(69)

\[
\hat{r}_{\min} l \leq \hat{R}_k \leq \hat{r}_{\max} l
\]  

(70)

for \( k \geq 0 \), such that the following conditions hold:

(1) The general nonlinear system given by (1) satisfies the observability rank condition for every vector \( x_k \in \mathbb{R}^r \).

(2) Nonlinear functions \( f(\cdot) \) and \( h(\cdot) \) are twice continuously differentiable with respect to their independent variables in \( x \), and also \( |\partial f / \partial x| \neq 0 \) holds for every vector \( x_k \in \mathbb{R}^r \).

Then the estimation error \( \tilde{x}_k \) given by (10) is exponentially bounded in the mean square.

Proof: In the proof of Theorem 4.1 relies on the use of Theorem 3.1. It is shown here that the inequalities (68)-(70) and the stated Conditions 1 and 2 in Theorem 4.1 together with the observability results given by Lemmas 4.1 and 4.3 imply the Conditions (34)-(36) in Theorem 3.1. Namely, it can be seen at once that (34)-(36) and (68)-(70) coincide; in other words, inequalities (34)-(36) are satisfied if inequalities (68)-(70) are.

Suppose functions \( f_i, h_i \) are the components of \( f \) and \( h \), respectively. Since \( f \) and \( h \) are twice differentiable for every independent variable in vector \( x \) according to Assumption 2 and \( \mathbb{R}^r \) is compact, the Hessian matrices of \( f_i \) and \( h_i \) are bounded with respect to the spectral norm of matrices. It is therefore that constants \( k_f \) and \( k_h \) are given by

\[ k_f = \max_{x_{\in R^d}} \| \text{Hess } f(x) \|, \quad k_h = \max_{x_{\in R^d}} \| \text{Hess } h(x) \| \]  
(71)

As far as the remaining conditions of Theorem 3.1 are concerned, it is sufficient to ensure these conditions hold one time-step in advance. It is important to notice that constants \( f_{\min}, f_{\max}, h_{\min}, h_{\max}, \lambda_{\min}, \lambda_{\max}, \tau_{\min}, \tau_{\max}, \rho_{\min}, \rho_{\max} \) in (29)-(33) can be chosen independently of the time \( k \).

This in turn means the boundedness of \( \tilde{x}_k \) and \( x_k \) implies the desired bounds on \( \Lambda_k, F_k, \Gamma_k, H_k \) and \( \hat{P}_k \), which are needed. Then from Theorem 3.1 the boundedness of \( \tilde{x}_k+1 \) for the next time step is readily obtained. By repeating this procedure again the bounds on \( \Lambda_{k+1}, F_{k+1}, \Gamma_{k+1}, H_{k+1} \) and \( \hat{P}_{k+1} \) are obtained, and therefore on \( \tilde{x}_{k+2} \) as well. Continuation of this proving strategy yields the desired result.

In order to establish the bounds on \( \Lambda_k, F_k, \Gamma_k, H_k, \) and \( \hat{P}_k \) note that the cases \( 0 \leq k < r \) and \( k \geq r \) can be treated separately. This is due to the fact that finite steps are needed to set up the uniform observability condition.

Firstly, the initial finite step cycle \( 0 \leq k < r \) is considered. By considering the boundedness of \( x_k, \tilde{x}_k, \) and of \( \tilde{x}_k, \Lambda_k, F_k, \Gamma_k, H_k, \) it follows that \( \hat{P}_k+1 > 0 \) if \( \hat{P}_k > 0 \). Now taking the minimum and maximum eigenvalue of \( \hat{P}_k \) and the maximum singular value of \( \Lambda_k, F_k, \Gamma_k, H_k \) for \( 0 \leq k < r \), the bounds (29)-(33) for \( 0 \leq k < r \) are derived.

Secondly, the case of the setup steps \( k \geq r \) is considered next. It should be noted that neither any eigenvalue of \( \hat{P}_k \) converges to zero nor any of the matrices \( \Lambda_k, F_k, \Gamma_k, H_k, \hat{P}_k \) diverges. The boundedness

\[ p_{\min} I \leq \hat{P}_k \leq p_{\max} I \]  
(72)

follows according to Lemma 4.3 by utilizing the boundedness of \( \dot{x}_k \) for \( r \leq k \leq k \) in a given region \( \| \dot{x}_k \| \leq \varepsilon, \varepsilon > 0 \).

Moreover, the norm boundedness of \( \Lambda_k, F_k, \Gamma_k, H_k \) also follows from the continuity of \( \partial f / \partial x \) and \( \partial h / \partial x \), the compactness of \( R^r \), and the fact that estimated samples \( \tilde{x}_k \) in \( R^r \). By means of these arguments, Theorem 3.1 is readily applied, which terminates the proof. \( \square \)

The obtained results of this section and of the preceding one showed the estimation error of the discrete-time Unscented Kalman Filter remains bounded provided that the general nonlinear system (1) satisfies the appropriate conditions without the requirements of: (i) a sufficiently small initial estimation error; and (ii) sufficiently weak noise. The latter precisely mark the major benefits of the new results reported in this study.

V. THE EXAMPLE: NUMERICAL AND SIMULATION RESULTS

To illustrate the significance of the derived theorems and the respective conditions, in this section the UKF is applied to a relevant example system. Evolution behavior of the error of the discrete-time unscented Kalman filter in both versions, the basic UKF and the modified MUKF, is then verified by numerical simulations. This section is written in two parts up. The first one is devoted to verify the theoretical result in Section III, while the second one is to verify the result in Section IV.

| TABLE I. INITIAL VALUES AND NOISE-WEIGHTING MATRICES FOR THE NUMERICAL SIMULATION |
|---------------------------------|-----------------|-----------------|-----------------|
|                                | Small initial error and small noise | Large noise | Large initial error |
| \( \tilde{x}_0 \)              | \( [0.5 \ 0.5]^T \)                      | \( [0.5 \ 0.5]^T \)                      | \( [2.3 \ 2.2]^T \)                      |
| \( G_k \)                      | \( \sqrt{10^{-5}} I \)                     | \( \sqrt{10^{-3}} I \)                     | \( \sqrt{10^{-5}} I \)                     |
| \( D_k \)                      | \( \sqrt{10} \)                              | \( \sqrt{10} \)                              | \( \sqrt{10} \)                              |
| \( \Delta Q_k \)               | \( \text{diag}[0.015^2 \ 0.02^2] \)       | \( \text{diag}[0.015^2 \ 0.02^2] \)       | \( \text{diag}[0.018^2 \ 0.4^2] \)       |

The following example system of the general nonlinear stochastic class (1) that is described with the model functions \( f(\cdot) \) and \( h(\cdot) \) as given

\[ f(x_k) = \begin{bmatrix} x_{k+1} + 0.2 \sigma_2 \sigma_3 \sigma_4 \sigma_5 \\ \frac{1}{1 + x_{k+1} \sigma_2} \end{bmatrix} \]  
(73)

\[ h(x_k) = \exp(-x_{k+1}) \]  
(74)

is considered. From (73) and (74) it is readily calculated:

\[ F_k = \frac{\partial f}{\partial x}(x_k) = \left[ \begin{array}{cc} 1 & 0 \\ -1 + 2 x_{k+1} \sigma_3 & 1 + \sigma_3 \sigma_4 \end{array} \right] \]  
(75)

\[ H_k = \frac{\partial h}{\partial x}(x_k) = \left[ \begin{array}{c} -\exp(-x_{k+1}) \\ 0 \end{array} \right] \]  
(76)

The following initial data have been chosen: \( Q_k = \text{I} \), \( R_k = 1/\tau \), \( P_0 = I_2 \), \( P_0 = I_2 \), \( x_0 = [0.8 \ 0.2]^T \). The sampling time chosen is \( \tau = 0.001 \), and the executing time steps are \( k = 10^4 \). Matrices \( G_k \) and \( D_k \) as well as the initial value \( \tilde{x}_0 \) have been chosen for each indicated particular case as shown in Table 1.

The numerical simulations results according to Theorem 3.1 of Section III are discussed first. In order to fulfill the assumption of Theorem 3.1 as shown in (29)-(36), the extra matrix \( \Delta Q_k = \text{diag}[0.015^2 \ 0.02^2] \) is designed by experiment and added in the UKF. Notice that deliberately a case with strong process noise and large initial error has been explored.

The relevant simulation results are depicted in Figures 2 and 3. In these figures, sample results for the unknown state \( \hat{x}_{2,k} \) and \( \hat{x}_{2,k} \) the estimated state as well as for the estimation error are plotted versus the discrete time.

The same simulation procedure is applied to this nonlinear model. There have been generated 10 trajectories and each of the nonlinear filters is performed for
10 times. The error standard deviations of the UKF and the 
EKF with different choices of $\Delta Q_k = \text{diag}(0.018^2, 0.4^2)$.

seen in Figures 1(c) and 2(c), the estimation error of EKF and 
UKF appear to be divergent. In contrast, the MUKF still can 
achieve the estimation error to remain well bounded.

It can be seen in Figures 1(a) and 2(a) that in the case of 
small initial error and small noise, the estimation error of 
the UKF and the MUKF is considerably smaller than that of 
the EKF. When the large noise is introduced to act (initial 
error kept unchanged), the estimation error of both the UKF 
and the MUKF remains bounded as simulation results in Figures 
1(b) and 2(b) show. For the case when large initial error is 
taken into consideration (noise is kept unchanged), as it can be 

The plots verify that provided a sufficiently large valued 
matrix $\Delta Q_k$ is chosen and added, the discrete-time MUKF 
appears to be rather robust. However, setting very large 
valued matrix $\Delta Q_k$ make the estimation error divergent for
the other two filters, as shown in Figures 1(c) and 2(c). In addition, the EKF appears to be an efficient estimator for this application, but its error standard deviation obviously.

For the treatment of the original nonlinear system case we turn to the theoretical results of Section IV. In order to fulfill the Assumption 1 in Theorem 4.1 the general nonlinear system has to satisfy the observability rank condition for every \( x_k \in \mathbb{R}^2 \). By using (60) and (78), (79) it is obtained

\[
U(x_k) = \begin{bmatrix}
-\exp(-\tilde{x}_{1,k}) & 0 \\
-\exp(-\tilde{x}_{1,k}) & -\exp(-\tilde{x}_{1,k})
\end{bmatrix}
\]

Therefore, the calculation yields \( \text{rank} U(x_k) = 2 \).

For \( x_k \in \mathbb{R}^2 \), Assumption 1 holds. With (73), (74), and (75) it can be easily checked that Assumption 2 is also fulfilled. The simulation results are depicted in Figures 1 and 2, where sample paths for the unknown state \( x_{2,k} \) and the estimated state \( \hat{x}_{2,k} \) as well as for the estimation error \( \tilde{x}_{2,k} \) are plotted versus \( k \). In the case of large measurement noise or large initial error, the estimation error remains bounded, which is due to the extra additive matrix \( \Delta Q_k \). Moreover, the EKF appears to be an efficient estimator for this particular application, but its error standard deviation is obviously large.

For \( x_k \in \mathbb{R}^2 \), Assumption 1 holds. With (73), (74), and (75) it can be easily checked that Assumption 2 is also fulfilled. The simulation results are depicted in Figures 1 and 2, where sample paths for the unknown state \( x_{2,k} \) and the estimated state \( \hat{x}_{2,k} \) as well as for the estimation error \( \tilde{x}_{2,k} \) are plotted versus \( k \). In the case of large measurement noise or large initial error, the estimation error remains bounded, which is due to the extra additive matrix \( \Delta Q_k \). Moreover, the EKF appears to be an efficient estimator for this particular application, but its error standard deviation is obviously large.

Apparently, the simulation results on the explored example confirm that the discrete-time UKF and the appropriate conditions for the stability of the discrete-time modified UKF are all effective. Thus the theory is supported by the application example of a nonlinear dynamic system.

VI. CONCLUDING REMARKS

The error dynamics behavior of the Unscented Kalman Filter when applied to estimation problems for nonlinear stochastic discrete-time systems of general type has been thoroughly investigated. It was shown in Section III that the estimation error is bounded in the mean square under certain conditions, which have been derived. According to some of the standard results on the boundedness of stochastic processes, the stability of the UKF can be ensured without the requirement of small initial estimation error by means of the appropriate choice of an additive positive definite matrix \( \Delta Q_k \).

This matrix, however, has to be designed by empirical tests; the only guidance is that it should be sufficiently large valued. Nonetheless, if this additional positive definite matrix \( \Delta Q_k \) is set too large valued then the standard deviations may becomes significant, which degrades the estimation quality. Therefore the design effort for matrix \( \Delta Q_k \) may be seen as a trade-off between the requirements for stability and for accuracy.

As shown in Section IV, the condition established in Section III can be reduced to a nonlinear observability rank condition of the plant model, which can be checked in advance. Considerably high performance of the UKF has been demonstrated under the worst initial conditions through numerical examples and simulations, sample results of which are given in the previous section.

REFERENCES


