SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2018, Vol. 61 042201:1–042201:16
https://doi.org/10.1007/s11432-017-9242-8

On extended state based Kalman filter design for a class of nonlinear time-varying uncertain systems

Wenyan BAI[1], Wenchao XUE[2,3*], Yi HUANG[2,3] & Haitao FANG[2,3]

[1]*Beijing Aerospace Automatic Control Institute, Beijing* 100039*, China;*
[2]*LSC, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China;*
[3]*School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, China*

Received 5 June 2017/Revised 4 August 2017/Accepted 22 September 2017/Published online 5 March 2018

Abstract This paper considers the filtering problem for a class of multi-input multi-output systems with nonlinear time-varying uncertain dynamics, random process and measurement noise. An extended state based Kalman filter, with the idea of timely estimating the unknown dynamics, is proposed for better robustness and higher estimation precision. The stability of the proposed filter is rigorously proved for nonlinear time-varying uncertain system with weaker stability condition than the extended Kalman filter, i.e., the initial estimation error, the uncertain dynamics and the noises are only required to be bounded rather than small enough. Moreover, quantitative precision of the proposed filter is theoretically evaluated. The proposed algorithm is proved to be the asymptotic unbiased minimum variance filter for constant uncertainty. The simulation results of some benchmark examples demonstrate the feasibility and effectiveness of the method.

Keywords Kalman filter, extended state observer, nonlinear time-varying uncertain system, unbiased minimum variance filter, active disturbance rejection control

Citation Bai W Y, Xue W C, Huang Y, et al. On extended state based Kalman filter design for a class of nonlinear time-varying uncertain systems. Sci China Inf Sci, 2018, 61(4): 042201, https://doi.org/10.1007/s11432-017-9242-8

1 Introduction

As is well known, Kalman filter (KF) has been widely used in many fields [1–3]. The original KF algorithm provides the minimum mean square error estimation for the accurate linear systems with Guass white noise. In the last decades, substantial development has been made on extending KF to deal with nonlinear uncertain systems [4, 5]. For nonlinear systems only perturbed by Gaussian noises, extended Kalman filter (EKF) is constructed by applying KF algorithm to the first-order linearized part of the nonlinear systems [6–8]. Unscented Kalman filter (UKF) captures posterior mean and covariance accurately to the third-order linearization for any nonlinearity by using a sampling approach [9]. For uncertain nonlinear system, robust filters (RF), including H_∞ filter, set valued filter (SVF), emphasize the issue of minimizing the estimation error under the deterministic/stochastic uncertainties of systems being in the worst-case bound [10, 11]. Substantial development of this issue has been made by several important literatures. Ref. [12] minimizes the mean square error according to the least favorable model with model perturbations limited to the bound parameterized by the τ-divergence family. Ref. [13] designs robust Kalman filter for linear systems with norm-bounded parameter uncertainty. Ref. [14]

* Corresponding author (email: wenchaoxue@amss.ac.cn)

minimizes the upper bound of the estimation error covariance for system with stochastic nonlinearities of zero means. Actually, linearization and worst-case bound optimization are two natural ideas to handle nonlinearity and uncertainty [15]. These two approaches have been proved to be effective in tackling many filtering problems [11, 16, 17]. Nevertheless, they still have the following limitations as in dealing with general nonlinear uncertain systems [9, 18].

(i) Linearization based filters may suffer from the instability issue in dealing with nonlinear system. Generally speaking, filters derived by local linearization at the current estimates have been developed based on an implicit assumption that the initial estimation error and the noise are sufficiently small to ensure the stability of the filter [7, 19]. Ref. [7] even gives the range of the initial estimation error and noise to ensure stability. Such an result is, unfortunately, very conservative and is hard to be satisfied in practical engineering system due primarily to the various external disturbances and the finite accuracy of the sensor during filter implementation. Thus the performance of filters based on linearization can be extremely fragile to initial estimation error and noise, which is shown in [7]. In conclusion, stability of linearization based filters is still an open problem for nonlinear system.

(ii) The optimization result for the worst-case bound of uncertainties may be conservative. RF usually has the idea of minimizing certain index based on the worst-case bound of uncertainties. This inevitably requires the bound of the uncertainty to be small enough otherwise the optimization result may be conservative. However, this assumption is critical when the uncertainty includes the unmodelled dynamics as well as the linearization error mentioned above. In this case, the uncertainty depends on the estimation error. Thus even the boundness of uncertainty should be rigorously proved, not to mention assuming small enough in advance [16, 20].

The above problems give the primary motivation to design resilient filters capable of dealing with strong nonlinearity and large uncertainty. Extended state observer (ESO) treats the nonlinear uncertainty, no matter how complex the uncertain dynamics are, as a time-varying signal to be estimated timely in order to correct the estimation error of the state. The idea of extended state is nature since compared to the noise, the nonlinear uncertain dynamics can be viewed as a relatively slow-varying signal which would be estimated in real time. Thus, the effectiveness of extended state design has been shown in lots of successfully applications, involving the flight systems [21–23], robotic systems [24], motor systems [25], MEMS systems [26] and so on [27]. However, parameters tuning for ESO is still an open problem, especially when the measurement noise is concerned, because higher gain leads to faster tracking, but also means the worse polluted by the noise [28]. In conclusion, to tackle the problems induced by the frame of linearization [7, 9] and the frame of worst-case optimization [12, 13], we will utilize the extended state design to handle the nonlinearity and uncertainty. In addition, we will also optimize the gain of the ESO timely to generate the extended state based Kalman filter (ESKF). The main contributions of ESKF can be highlighted as follow.

(i) The gain of ESO is optimized timely to improve the estimation performance for system with both uncertain dynamics and stochastic noise.

(ii) Stability of the filter is guaranteed for nonlinear system with strong nonlinearity, large initial estimation error and large noise.

(iii) Better robustness against large scale of uncertainty is guaranteed by actively estimating the nonlinear uncertainty.

In fact, the effectiveness of the extended state design in filters has been shown by [29, 30] for a class of time-invariant uncertain systems. Nevertheless, the systems considered therein are limited into time-invariant. Also the stability and the performance of the corresponding filters have not been discussed for time-varying systems. In this paper, we construct ESKF for a more general class of nonlinear time-varying uncertain systems. With the aim of weakening the stability condition and improving the estimation precision for the existing filters despite nonlinear uncertainties and noise terms, we prove that the estimation error of ESKF is bounded in mean square despite nonlinear uncertainties, large noise and large initial estimation error. In addition, ESKF's estimation precision can be timely evaluated by its parameters, and it is proved to be the asymptotically unbiased minimum variance estimation under certain conditions.

The paper is organized as follows. In Section 2, ESKF for a class of nonlinear uncertain systems is proposed. In Section 3, the properties of ESKF are discussed. In Section 4, the effectiveness of ESKF is shown by two classical examples. The concluding remark is given in Section 5.

2 Problem formulation and ESKF design

Consider the following class of nonlinear time-varying uncertain systems:

$$\begin{cases} X_{k+1} = \bar{A}_k X_k + \bar{B}_k F(X_k, k) + w_k, \\ Y_k = \bar{C}_k X_k + n_k, \end{cases} \quad k = 0, 1, \ldots, \tag{1}$$

where $X_k \in \mathbb{R}^n$ is the state, $\bar{A}_k$, $\bar{B}_k$ and $\bar{C}_k$ are known time-varying matrixes with $\bar{A}_k \in \mathbb{R}^{n\times n}$, $\bar{B}_k \in \mathbb{R}^{n\times l}$, $\bar{C}_k \in \mathbb{R}^{m\times n}$. $F(X_k, k) \in \mathbb{R}^l$ is the nonlinear uncertain dynamics in the system (1), and its nominal model is the known function $\overline{F}(X_k, k)$. $w_k \in \mathbb{R}^n$ and $n_k \in \mathbb{R}^m$ are the process noise and measurement noise, respectively. $Y_k \in \mathbb{R}^m$ is the measurement output.

Remark 1. $F(X_k, k)$ is usually referred to the "total disturbance" lumping both internal uncertain dynamics and external disturbance, such as the unknown parameter variations, the unmodeled dynamics and the discretization error [31, 32].

In model (1), the uncertain dynamics are divided into three parts, i.e., the known linear part $\bar{A}_k X_k$, the slowly time-varying nonlinear uncertain dynamics $F(X_k, k)$ and the noise (w_k, n_k) which may have high-frequency changes. We suggest using different methods to deal with different kinds of uncertainties. $F(X_k, k)$ is treated as an extended state to be estimated as well as compensated for, and (w_k, n_k) is attenuated by the optimization technique of KF. Therefore, $F(X_k, k)$ is treated as an extended state and system (1) can be equivalently transformed to

$$\begin{cases} \begin{bmatrix} X_{k+1} \\ F_{k+1} \end{bmatrix} = A_k \begin{bmatrix} X_k \\ F_k \end{bmatrix} + B_k G_k + \begin{bmatrix} w_k \\ 0 \end{bmatrix}, \\ Y_k = C_k \begin{bmatrix} X_k \\ F_k \end{bmatrix} + n_k, \end{cases} \tag{2}$$

where

$$F_k \triangleq F(X_k, k), \quad G_k = F_{k+1} - F_k, \quad A_k = \begin{bmatrix} \bar{A}_k & \bar{B}_k \\ 0 & I \end{bmatrix}, \quad B_k = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad C_k = \begin{bmatrix} \bar{C}_k & 0 \end{bmatrix}. \tag{3}$$

Next, we will construct ESKF for the system (2). To begin with, the following assumptions are given.

A1. (A_k, C_k) is uniformly observable.

A2. $\{w_k\}_0^\infty$ and $\{n_k\}_0^\infty$ are uncorrelated zero-mean Gaussian random sequences and

$$E(n_k {n_k}^{\mathrm{T}}) \leqslant R_k, \quad E(w_k {w_k}^{\mathrm{T}}) \leqslant S_k, \tag{4}$$

where $\{R_k\}_{k=0}^\infty$ and $\{S_k\}_{k=0}^\infty$ are known and uniformly bounded. In addition, $\{w_k\}_0^\infty$, $\{n_k\}_0^\infty$ and X_0 are mutually independent.

A3.

$$E\left(\begin{bmatrix} X_0 - \hat{X}_0 \\ F_0 - \hat{F}_0 \end{bmatrix} \begin{bmatrix} X_0 - \hat{X}_0 \\ F_0 - \hat{F}_0 \end{bmatrix}^{\mathrm{T}} \right) \leqslant P_0, \tag{5}$$

where $\hat{X}_0$ is the estimate of X_0, $\hat{F}_0 \triangleq \bar{F}(\hat{X}_0, 0)$, and P_0 is a known constant matrix.

A4.

$$E(G_{k,i}^2) \leqslant \overline{q}_{k,i}, \quad i = 1, 2, \ldots, \quad \forall k \geqslant 0, \tag{6}$$

where $\{\overline{q}_{k,i}\}_{k=0}^\infty$ is known and uniformly bounded[1)].

1) In this paper, if a_k is a vector, then $a_{k,i}$ denotes its i-th element.

The assumption A1 is an important and fundamental assumption to ensure the stability of KF typed algorithm [3, 33]. From (3) we know that A_k and C_k are determined by the matrices of the original system (1), which means the observability assumption of (A_k, C_k) is actually added on the structure of the original system. The relationship between the observability of (A_k, C_k) and $(\bar{A}_k, \bar{C}_k)$ can be demonstrated as the following lemma.

Lemma 1. For systems (1) and (2), (A_k, C_k) is uniformly observable if and only if both $(\bar{A}_k, \bar{C}_k)$ is uniformly observable and rank$\left[\begin{smallmatrix} I_{n\times n}-\bar{A}_k & -\bar{B}_k \\ \bar{C}_k & 0 \end{smallmatrix}\right] = n + l$, for all $k = 0, 1, 2, \ldots$.

The proof of Lemma 1 is in Appendix A. From Lemma 1, the uniformly observability of (A_k, C_k) equals to both the uniformly observability of $(\bar{A}_k, \bar{C}_k)$ and rank$\left[\begin{smallmatrix} I_{n\times n}-\bar{A}_k & -\bar{B}_k \\ \bar{C}_k & 0 \end{smallmatrix}\right] = n+l,\ k = 0, 1, 2, \ldots$. However, compared with the latter conditions, uniformly observability of (A_k, C_k) seems more brief and easier to be verified. So we choose A1 as the first assumption.

Remark 2. The assumptions A3 and A4 mean that the estimation error for the initial states, the estimation error for the initial uncertain dynamics and the varying of the uncertain function $F(\cdot)$ between neighborhood time steps are bounded in mean square. Obviously, A3 and A4 are reasonable for most practical plants. $\overline{q}_k$ represents the size of the varying of the nonlinear uncertainty $F(\cdot)$ [34]. Besides, the upper bounds P_0 and $\overline{q}_k$ can be chosen according to the priori information of the sensors and physical limitations on the practical systems.

Then, we design ESO [31] based on the extended model (2)

$$\begin{bmatrix} \hat{X}_{k+1} \\ \hat{F}_{k+1} \end{bmatrix} = A_k \begin{bmatrix} \hat{X}_k \\ \hat{F}_k \end{bmatrix} + B_k \hat{G}_k - K_k \left(Y_k - C_k \begin{bmatrix} \hat{X}_k \\ \hat{F}_k \end{bmatrix} \right). \tag{7}$$

Here, $\hat{G}_k$, the estimate of G_k, is used to correct the estimation error of the state and the uncertainty by making full use of the model information. Thus, we use the nominal model of G_k as

$$\bar{G}_k = \bar{F}(X_{k+1}, k + 1) - \bar{F}(X_k, k). \tag{8}$$

Then the estimate of $\bar{G}_k$ is denoted as

$$\hat{\bar{G}}_k = \bar{F}(\bar{A}_k \hat{X}_k + \bar{B}_k \hat{F}_k, k + 1) - \bar{F}(\hat{X}_k, k). \tag{9}$$

According to the estimate of the nominal model $\bar{G}_k$, $\hat{G}_k$ is designed as

$$\hat{G}_{k,i} = \text{sat}\left(\hat{\bar{G}}_{k,i}, \sqrt{\overline{q}_{k,i}}\right), \quad i = 1, 2, \ldots, l, \tag{10}$$

where sat$(\cdot)$ is the saturation function defined by sat$(f, b) = \max\{\min\{f, b\}, -b\},\ b > 0$. Here, the saturation function sat$(\cdot)$ is used to ensure the boundedness of $\hat{G}_{k,i}$.

For the system (1) with uncertainty and noise, the tuning of K_k becomes a tradeoff between the disturbance rejection and the noise sensitivity, because higher K_k leads to faster tracking, but also means the worse polluted by the noise [28]. Thus, different from ESO, K_k is no longer static and manually tuned here. Instead, we will optimize K_k to make the mean square estimation error minimal at each step.

Denote the estimation error of ESKF as $e_k = \left[\begin{smallmatrix} X_k \\ F_k \end{smallmatrix}\right] - \left[\begin{smallmatrix} \hat{X}_k \\ \hat{F}_k \end{smallmatrix}\right]$. From (2) and (7), we can obtain that e_k satisfies

$$e_{k+1} = (A_k + K_k C_k) e_k + K_k n_k + B_k (G_k - \hat{G}_k) + \begin{bmatrix} w_k^{\mathrm{T}} & 0 \end{bmatrix}^{\mathrm{T}}. \tag{11}$$

Denote $\tilde{G}_k = B_k(G_k - \hat{G}_k)$. Since n_k, w_k and e_k are mutually independent, the mean square error of ESKF satisfies

$$\begin{aligned} E(e_{k+1} e_{k+1}^{\mathrm{T}}) &= (A_k + K_k C_k)\, E(e_k e_k^{\mathrm{T}})\, (A_k + K_k C_k)^{\mathrm{T}} + K_k E(n_k n_k^{\mathrm{T}}) K_k{}^{\mathrm{T}} + E(\tilde{G}_k \tilde{G}_k^{\mathrm{T}}) \\ &\quad + \begin{bmatrix} E(w_k w_k^{\mathrm{T}}) & 0_{n\times l} \\ 0_{l\times n} & 0_{l\times l} \end{bmatrix} + E((A_k + K_k C_k)\, e_k \tilde{G}_k^{\mathrm{T}}) + E(\tilde{G}_k e_k^{\mathrm{T}}\, (A_k + K_k C_k)^{\mathrm{T}}). \end{aligned} \tag{12}$$

According to the Young's inequality for the matrices case and A4, we can get

$$\begin{aligned}(G_k-\hat{G}_k)(G_k-\hat{G}_k)^{\mathrm{T}} &\leqslant 2G_kG_k^{\mathrm{T}}+2\hat{G}_k\hat{G}_k^{\mathrm{T}}\\ &\leqslant 2l\mathrm{diag}([G_{k,1}^2\ G_{k,2}^2\ \cdots\ G_{k,l}^2])+2l\mathrm{diag}([\hat{G}_{k,1}^2\ \hat{G}_{k,2}^2\ \cdots\ \hat{G}_{k,l}^2])\\ &\leqslant 4l\mathrm{diag}([\bar{q}_{k,1}\ \cdots\ \bar{q}_{k,l}]).\end{aligned}\tag{13}$$

Design $Q_{1,k}=\begin{bmatrix}0_{n\times n} & 0_{n\times l}\\ 0_{l\times n} & 4\overline{Q}_k\end{bmatrix}$, where $\overline{Q}_k\triangleq l\cdot\mathrm{diag}([\overline{q}_{k,1}\ \overline{q}_{k,2}\ \ldots\ \overline{q}_{k,l}])$. Then, it is easy to know that

$$E((G_k-\hat{G}_k)(G_k-\hat{G}_k)^{\mathrm{T}})\leqslant 4\overline{Q}_k,\quad E(\tilde{G}_k\tilde{G}_k^{\mathrm{T}})\leqslant Q_{1,k}.\tag{14}$$

Moreover, since the noise related to F_k is correlated to the noise affecting X_k, the corresponding term of e_k and $\tilde{G}_k$ cannot be ignored. Based on the Young's inequality for the matrices case, the last two terms of (12) have the upper bound

$$\begin{aligned}&(A_k+K_kC_k)\,e_k\tilde{G}_k^{\mathrm{T}}+\tilde{G}_k{e_k}^{\mathrm{T}}\,(A_k+K_kC_k)^{\mathrm{T}}\\ &\leqslant\theta\,(A_k+K_kC_k)\,e_ke_k^{\mathrm{T}}\,(A_k+K_kC_k)^{\mathrm{T}}+\frac{1}{\theta}\tilde{G}_k\tilde{G}_k^{\mathrm{T}},\quad\forall\theta>0.\end{aligned}\tag{15}$$

Ideally, the equality holds if and only if

$$\theta\,(A_k+K_kC_k)\,e_k=\tilde{G}_k.\tag{16}$$

However, Eq. (16) usually cannot be achieved for each $k\geqslant 0$. Hence we consider the initial time $k=0$. It can be verified that the equality of (16) with $k=0$ has the necessary condition

$$\theta^2\,(A_k+K_0C_k)\,E(e_0e_0^{\mathrm{T}})(A_k+K_0C_k)^{\mathrm{T}}=E(\tilde{G}_0\tilde{G}_0^{\mathrm{T}})\leqslant Q_{1,0}.\tag{17}$$

For simplicity, we suggest $\theta=\sqrt{\frac{\mathrm{tr}(Q_{1,0})}{\mathrm{tr}(P_0)}}$ according to (5) and (17). We remark that θ is used here to decouple the cross terms of estimation error and the uncertainties.

Design

$$Q_{2,k}=\begin{bmatrix}S_k & 0_{n\times l}\\ 0_{l\times n} & 0_{l\times l}\end{bmatrix}.$$

Then, from A2 we have

$$\begin{bmatrix}E(w_kw_k^{\mathrm{T}}) & 0_{n\times l}\\ 0_{l\times n} & 0_{l\times l}\end{bmatrix}\leqslant Q_{2,k}.\tag{18}$$

Hence, according to (14)–(18) and A2, we have

$$E(e_{k+1}e_{k+1}^{\mathrm{T}})\leqslant(1+\theta)\,(A_k+K_kC_k)\,E(e_ke_k^{\mathrm{T}})\,(A_k+K_kC_k)^{\mathrm{T}}+K_kR_kK_k^{\mathrm{T}}+\left(1+\frac{1}{\theta}\right)Q_{1,k}+Q_{2,k}.\tag{19}$$

Let P_k satisfy the following iteration equation:

$$\begin{aligned}&P_{k+1}=(1+\theta)\,(A_k+K_kC_k)\,P_k\,(A_k+K_kC_k)^{\mathrm{T}}+K_kR_k{K_k}^{\mathrm{T}}+\left(1+\frac{1}{\theta}\right)Q_{1,k}+Q_{2,k},\\ &k=0,1,2,\ldots,\end{aligned}\tag{20}$$

with the initial value P_0.

Consequently, it follows from (19) and (20) that

$$E(e_ke_k^{\mathrm{T}})\leqslant P_k,\quad k=0,1,2,\ldots,\tag{21}$$

provided A3.

According to (21), P_k describes the upper bound of the mean square error $E(e_k e_k^{\mathrm{T}})$. To simplify the algorithm and to ensure the consistence property [35], we choose K_k at each time step to minimize P_k instead of directly minimizing $E(e_k e_k^{\mathrm{T}})$. The first two terms in the right hand of (20) are related to K_k, which motivates us to design K_k to minimize these two terms.

Define

$$K_k^* = \arg\min_{K_k} \{(1+\theta)\left(A_k + K_k C\right) P_k \left(A_k + K_k C\right)^{\mathrm{T}} + K_k R_k {K_k}^{\mathrm{T}}\}. \tag{22}$$

Since P_k is a positive semi-definite matrix, and R_k is a positive definite matrix, $CP_kC^{\mathrm{T}} + \frac{1}{1+\theta}R_k$ is positive definite. Thus, there is

$$\begin{aligned}
&(1+\theta)\left(A_k + K_k C\right) P_k \left(A_k + K_k C\right)^{\mathrm{T}} + K_k R_k {K_k}^{\mathrm{T}} \\
&= (1+\theta)\left(A_k P_k C^{\mathrm{T}}\left(C_k P_k C_k^{\mathrm{T}} + \frac{1}{1+\theta}R_k\right)^{-1} + K_k\right) \\
&\quad \cdot \left(C_k P_k C_k^{\mathrm{T}} + \frac{1}{1+\theta}R_k\right)\left(A_k P_k C_k^{\mathrm{T}}\left(C_k P_k C_k^{\mathrm{T}} + \frac{1}{1+\theta}R_k\right)^{-1} + K_k\right)^{\mathrm{T}} \\
&\quad - (1+\theta) A_k P_k C_k^{\mathrm{T}}\left(C_k P_k C_k^{\mathrm{T}} + \frac{1}{1+\theta}R_k\right)^{-1} C_k P_k A_k^{\mathrm{T}} + (1+\theta) A_k P_k A_k^{\mathrm{T}}.
\end{aligned} \tag{23}$$

Hence, it is easy to know

$$K_k^* = -A_k P_k C_k^{\mathrm{T}}\left(C_k P_k C_k^{\mathrm{T}} + \frac{1}{1+\theta}R_k\right)^{-1}. \tag{24}$$

Since then, the gain matrix K_k of the estimator is optimized timely based on the prior information A2–A4.

As a consequence, ESKF, which combines the advantages of ESO and KF, is designed as follows[2]:

$$\begin{bmatrix} \hat{X}_{k+1} \\ \hat{F}_{k+1} \end{bmatrix} = A_k \begin{bmatrix} \hat{X}_k \\ \hat{F}_k \end{bmatrix} + B_k \hat{G}_k - K_k \left(Y_k - C_k \begin{bmatrix} \hat{X}_k \\ \hat{F}_k \end{bmatrix}\right), \tag{25}$$

$$K_k = -A_k P_k C_k^{\mathrm{T}}\left(C_k P_k C_k^{\mathrm{T}} + \frac{1}{1+\theta}R_k\right)^{-1}, \tag{26}$$

$$P_{k+1} = (1+\theta)\left(A_k + K_k C_k\right) P_k \left(A_k + K_k C_k\right)^{\mathrm{T}} + K_k R_k {K_k}^{\mathrm{T}} + \left(1 + \frac{1}{\theta}\right) Q_{1,k} + Q_{2,k}, \tag{27}$$

$$Q_{1,k} = \begin{bmatrix} 0_{n\times n} & 0_{n\times l} \\ 0_{l\times n} & 4\overline{Q}_k \end{bmatrix}, \quad Q_{2,k} = \begin{bmatrix} S_k & 0_{n\times l} \\ 0_{l\times n} & 0_{l\times l} \end{bmatrix}, \quad \overline{Q}_k = l \cdot \mathrm{diag}([\overline{q}_{k,1}\ \overline{q}_{k,2}\ \ldots\ \overline{q}_{k,l}]), \tag{28}$$

where

$$\theta = \sqrt{\frac{\mathrm{tr}(Q_{1,0})}{\mathrm{tr}(P_0)}}, \quad \hat{G}_{k,i} \triangleq \mathrm{sat}\left(\bar{G}_{k,i}, \sqrt{\overline{q}_{k,i}}\right), \quad \bar{G}_k \triangleq \bar{G}(\hat{X}_k, k), \quad i = 1, 2, \ldots, l, \tag{29}$$

and $\mathrm{sat}(\cdot)$ is the saturation function defined by $\mathrm{sat}(f, b) = \max\{\min\{f, b\}, -b\},\ b > 0$.

According to (25)–(28), ESKF can be viewed as a novel KF typed filter for nonlinear time-varying uncertain systems. Moreover, ESKF and traditional KF-type filters have the following differences.

(i) The traditional KF-type filters usually approximate the real model as accurately as possible by using higher order linearization, while ESKF treats the nonlinear uncertainty as a time-varying signal to be estimated and compensated no matter how complex the model is. Thus ESKF avoids the linearization which may cause divergency.

(ii) Compared with RF, ESKF uses the bound for the varying rate of uncertain dynamics instead of the bound for uncertain dynamics itself to solve the optimization. Thus, RF usually needs the uncertainty to

2) In this paper, if a is a vector, then $\mathrm{diag}(a)$ is the matrix whose i-th diagonal element is a_i and off-diagonal elements are all 0.

be small enough to guarantee satisfactory performance while ESKF only assumes the nonlinear uncertain dynamics to be slow-varying.

(iii) Most existing filtering methods can only deal with single type of disturbance, while an actual filtering problem usually suffers from multiple disturbances. In ESKF, the disturbances are classified into two types, i.e., slow-varying uncertain dynamics and sharp-changing noise. Also, ESKF suggests tackling these two kinds of disturbances in different ways.

The properties of ESKF (25)–(28) will be discussed in Section 3.

3 The properties of ESKF

Theorem 1 (Stability)**.** If A1–A4 hold, then the estimation error of ESKF, $e_k = \begin{bmatrix} X_k \\ F_k \end{bmatrix} - \begin{bmatrix} \hat{X}_k \\ \hat{F}_k \end{bmatrix}$, satisfies

$$E(e_k e_k^{\mathrm{T}}) \leqslant P_k, \quad \forall k \geqslant 0, \tag{30}$$

and $\{P_k\}_{k=0}^{\infty}$ is uniformly bounded. Then, there exists a positive matrix P^* such that

$$P_k \leqslant P^*, \quad \forall k \geqslant 0. \tag{31}$$

The proof of Theorem 1 is in Appendix B. Theorem 1 demonstrates that ESKF (25)–(28) has the following properties.

(i) The estimation error of ESKF is bounded in mean square despite the nonlinear unknown dynamics, the process and measurement noise. Reif et al. [7] indicated that the stability of EKF can only be guaranteed for true nonlinear model under small initial estimation error and small disturbing noise. ESKF overcomes these constraints by augmenting the filter with an extended state to actively estimate the nonlinear uncertain part in the system (1). Therefore, ESKF provides a novel frame to handle large initial estimation error, nonlinear unknown dynamics as well as noises.

(ii) The estimation error, which is unavailable in practice, can be timely evaluated by the parameter P_k in mean square. For linear systems, the consistence property [9] is ensured, which means that P_k in KF exactly equals to the covariance of estimation error. However, in EKF, the relationship between P_k and the covariance of estimation error is vague, which may lead to unreliable filter gain and the possibility of divergence. Therefore, P_k in ESKF contains much more important information for evaluating the filtering precision.

As P_k is an upper bound in mean square for the estimation error, we will investigate the relationship between P_k and the tunable parameters $(P_0, Q_{1,k}, Q_{2,k}, R_k)$ in Lemma 2.

Lemma 2. Let P_k^* and P_k^{**} denote the P_k under the situations of $(P_0^*, Q_{1,k}^*, Q_{2,k}^*, R_k^*)$ and $(P_0^{**}, Q_{1,k}^{**}, Q_{2,k}^{**}, R_k^{**})$ in (25)–(28), respectively. If

$$P_0{}^* < P_0{}^{**}, \quad Q_{1,k}^* < Q_{1,k}^{**}, \quad Q_{2,k}^* < Q_{2,k}^{**}, \quad R_k^* < R_k^{**}, \tag{32}$$

then

$$P_k^* < P_k^{**}, \quad \forall k \geqslant 0. \tag{33}$$

The proof of Lemma 2 is in Appendix C. Lemma 2 illustrates the relationship between P_k and the tunable parameters $(P_0, Q_{1,k}, Q_{2,k}, R_k)$, that is, smaller $(P_0, Q_{1,k}, Q_{2,k}, R_k)$ satisfying A1–A4 mean smaller P_k. Obviously, Lemma 2 implies that $(P_0, Q_{1,k}, Q_{2,k}, R_k)$ are suggested to be tuned as small as possible so as to achieve better filtering performance.

Theorem 2 further shows ESKF is an asymptotically unbiased minimum variance estimation under certain conditions.

Theorem 2 (Optimality)**.** If $F(X_k, k) = F_0$, $\bar{A}_k = \bar{A}$, $\bar{B}_k = \bar{B}$, $\bar{C}_k = \bar{C}$, $R_k = R$, $S_k = S$, $\forall k \geqslant 0$, then ESKF (25)–(28) with $\overline{Q}_k = 0$ is an asymptotically unbiased minimum variance estimation for $[X_k^{\mathrm{T}} \ F_k^{\mathrm{T}}]^{\mathrm{T}}$ among all the functions of $\{Y_i\}_{i=1}^{k-1}$.

Table 1 Initial values and noise weighting matrices for Example 1

	$\hat{z}_0$	G_k	D_k
Small initial error and small noise	$[0.5\ 0.5]^{\mathrm{T}}$	$\sqrt{10^{-5}}I$	$\sqrt{10}$
Small initial error and large noise	$[0.5\ 0.5]^{\mathrm{T}}$	$\sqrt{10^{-5}}I$	$\sqrt{10^3}$
Large initial error and small noise	$[1.5\ 1]^{\mathrm{T}}$	$\sqrt{10^{-5}}I$	$\sqrt{10}$

The proof of Theorem 2 is in Appendix D. Theorem 2 shows that ESKF has some optimality in the sense of minimum variance if the uncertain term $F(X_k, k)$ is constant, even though it can be arbitrarily large and unknown. Note that the $F(X_k, k)$ in the physical plants is usually slowly time-varying dynamics, then it is not to hard to know that ESKF performs well in practical systems, as shown in the examples of Section 4.

Next, the effectiveness of ESKF for nonlinear filtering problem will be shown by two systems which have been studied with EKF and the traditional SVF, respectively.

4 Case studies

4.1 Example 1

Reif et al. [7] has shown that EKF may quickly diverge without the conditions of sufficient small initial estimation error or sufficient small noise. We apply ESKF to this example to illustrate that ESKF can weaken the stability conditions on the initial error and noise and guarantee the stability for nonlinear systems.

The system in [7] is given as follows:

$$\begin{cases} z_{k+1} = \bar{A}_k z_k + \bar{B}_k F_k + G_k w_k, \quad z_0 = [0.8\ 0.2]^{\mathrm{T}}, \\ y_k = \bar{C}_k z_k + D_k v_k, \end{cases} \tag{34}$$

where $z_k = \left[\begin{smallmatrix} z_{1,k} \\ z_{2,k} \end{smallmatrix}\right]$ is the state,

$$\bar{A}_k = \begin{bmatrix} 1 & \tau \\ -\tau & 1-\tau \end{bmatrix}, \quad \bar{B}_k = \begin{bmatrix} 0 \\ \tau \end{bmatrix}, \quad \bar{C}_k = [1\ 0], \quad F_k = z_{2,k}(z_{1,k}^2 + z_{2,k}^2), \tag{35}$$

y_k is the measurement, and w_k and v_k are uncorrelated zero-mean white noise processes with identity covariance. The sampling time is $\tau = 10^{-3}$, executing $k = 2 \times 10^4$ steps. The matrices G_k and D_k as well as the initial value $\hat{z}_0$ are shown as Table 1, which are the same as those in [7].

We treat F_k as an extended state and design ESKF with parameters

$$\hat{F}_0 = 0, \quad R_k = D_k^2, \quad \overline{Q}_k = 3 \times 10^{-5}, \quad Q_{2,k} = \begin{bmatrix} G_k^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad P_0 = I_{3\times 3}. \tag{36}$$

Figure 1 illustrates the paths for the estimation error $(\hat{z}_{2,k} - z_{2,k})$ of ESKF and EKF. It can be seen from Figure 1 that in the case of large measurement noise or large initial error, the estimation error of EKF is quickly divergent while the estimation error of ESKF is bounded. Thus ESKF releases the conditions of small enough initial error and disturbing noise, and shows low sensitivity to the initial error and noise.

4.2 Example 2

In order to further demonstrate the ability of ESKF in dealing with model uncertainties, we use ESKF and the traditional SVF [20, 36, 37] in maneuvering targets tracking and compare their filtering performances

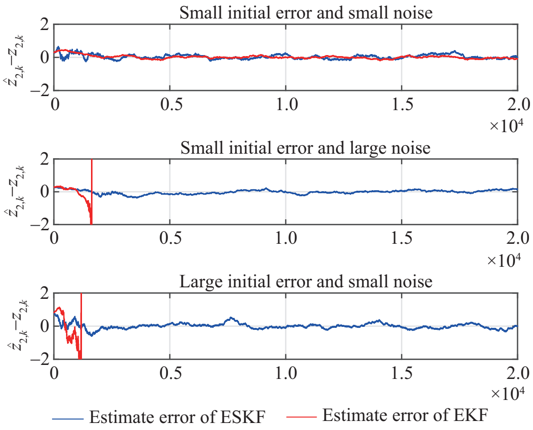

Figure 1 (Color online) The estimation error of ESKF and EKF (Example 1).

for the following target state-space model with unknown maneuvering [38]:

$$\begin{cases} \dot{S}(t) = V(t), \\ \dot{V}(t) = -2\Omega V(t) - \Omega^2 S(t) - \dfrac{\mu}{\|S(t)\|^3} S(t) + \Delta_F(t), \\ Y_k = \bar{C}_k S_k + v_k, \end{cases} \tag{37}$$

where $S(t)$ and $V(t)$ are the target position and velocity vectors. The terms $2\Omega V(t)$ and $\Omega^2 S(t)$ represent the accelerations due to the Coriolis and centrifugal forces respectively, where $\Omega = \left[\begin{smallmatrix} 0 & -\omega_e & 0 \\ \omega_e & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right]$. $\omega_e = 7.292115 \times 10^{-5}$ is the Earth rotation rate. $\frac{\mu}{\|S(t)\|^3} S(t)$ represents the gravitational acceleration, where $\mu = 3.986005 \times 10^{14}$ is the Earth's gravitational constant. $\Delta_F(t)$ is the acceleration induced by unknown maneuvering force or external disturbance.

Y_k is the sampled measurement output and the sampling time is $h = 0.01$ s. k denotes the integer.

$$\bar{C}_k = \begin{bmatrix} \mathrm{e}^{-9\times10^{-8}} & 0 & 0 \\ 0 & \mathrm{e}^{-9\times10^{-8}} & 0 \\ 0 & 0 & \mathrm{e}^{-4.5\times10^{-8}} \end{bmatrix}.$$

$\{v_k\}_0^\infty$ is uncorrelated Gaussian random sequence with its variance matrix being

$$\sigma_v^2 = \begin{bmatrix} 8\times10^3 & 0 & 0 \\ 0 & 4.5\times10^3 & 0 \\ 0 & 0 & 4.5\times10^3 \end{bmatrix}.$$

The targets tracking problem has been identified by a number of authors [39, 40], as being particularly stressful for filters and trackers because of the strong nonlinearities exhibited by the gravity. Besides, the tracking problem becomes more difficult for the existence of unknown acceleration $\Delta_F(t)$.

We treat the sum of the nonlinearity $\frac{\mu}{\|S(t)\|^3} S(t)$ and the uncertainty $\Delta_F(t)$ as the "total disturbance" $F(t)$, which satisfies

$$F(t) = -\frac{\mu}{\|S(t)\|^3} S(t) + \Delta_F(t).$$

Thus the nominal model (known information) of $F(t)$ is

$$\bar{F}(t) = \frac{\mu}{\|S(t)\|^3} S(t).$$

Denote the state X_k as $X_k \triangleq \begin{bmatrix} S(kh) \\ V(kh) \end{bmatrix}$. The model (37) is discretized with sampling time h in order to get the discrete form (1) with

$$\bar{A}_k = I_{6\times 6} + h \begin{bmatrix} 0_{3\times 3} & I_{3\times 3} \\ -\Omega^2 & -2\Omega \end{bmatrix}, \quad \bar{B}_k = \begin{bmatrix} 0_{3\times 3} \\ hI_{3\times 3} \end{bmatrix}.$$

To compare the robustness of ESKF and traditional SVF against different uncertainties, the simulations are carried out for the following three cases of the unknown acceleration Δ_F:

$$C1: \Delta_F(t) = 0.1\sin(0.02t), \quad C2: \Delta_F(t) = 1.5\sin(0.02t), \quad C3: \Delta_F(t) = 0.1t.$$

According to (25)–(28), ESKF is designed with

$$\hat{X}_0 = \begin{bmatrix} \bar{C}_0^{-1}Y_0 \\ 0_{3\times 1} \end{bmatrix}, \quad \hat{F}_0 = \frac{\mu}{\|\bar{C}_0^{-1}Y_0\|^3}\bar{C}_0^{-1}Y_0, \quad P_0 = \begin{bmatrix} \bar{C}_0^{-1}\sigma_v^2\bar{C}_0^{-1} & & \\ & v_{\max}^2 I_{3\times 3} & \\ & & I_{3\times 3} \end{bmatrix}, \tag{38}$$

$$R_k = \sigma_v^2, \quad Q_{1,k} = \begin{bmatrix} 0_{6\times 6} & 0_{6\times 3} \\ 0_{3\times 6} & 4\times 10^{-7} I_{3\times 3} \end{bmatrix}^2. \tag{39}$$

The traditional SVF is designed as [36]

$$\hat{X}_k = \bar{X}_k + K_k\left(y_k - \bar{C}_k\bar{X}_k\right), \tag{40}$$

$$\bar{X}_{k+1} = \bar{A}_k\hat{X}_k + \bar{B}_k\bar{F}(\hat{X}_k), \tag{41}$$

$$K_k = \tilde{w}\bar{P}_k\bar{C}_k^{\mathrm{T}}\left(\tilde{w}\bar{C}_k\bar{P}_k\bar{C}_k^{\mathrm{T}} + R_k\right)^{-1}, \tag{42}$$

$$P_k = (I - K_k\bar{C}_k^{\mathrm{T}})\bar{P}_k, \tag{43}$$

$$\bar{P}_{k+1} = \frac{1}{1-\rho_k}J_kP_kJ_k^{\mathrm{T}} + \frac{1}{\rho_k}Q_k, \tag{44}$$

$$\bar{X}_0 = \begin{bmatrix} \bar{C}_0^{-1}Y_0 \\ 0 \end{bmatrix}, \quad \bar{P}_0 = \begin{bmatrix} (\bar{C}_0^{-1}\sigma_v)^2 & \\ & v_{\max}^2 I_{3\times 3} \end{bmatrix}, \quad R_k = \sigma_v^2, \tag{45}$$

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{30}I_{3\times 3} \end{bmatrix}, \quad \rho_k = \frac{\sqrt{\mathrm{tr}(Q_k)}}{\sqrt{\mathrm{tr}(Q_k)} + \sqrt{\mathrm{tr}(J_kP_kJ_k^{\mathrm{T}})}}, \quad \tilde{w} = 0.05, \tag{46}$$

where

$$J_k \triangleq \left.\frac{\partial(\bar{A}_kX + \bar{B}_kF(X))}{\partial X}\right|_{\bar{X}_k} = I_{6\times 6} + h\begin{bmatrix} 0_{3\times 3} & I_{3\times 3} \\ -\left(\frac{\mu}{\|\bar{S}_k\|^3}I_{3\times 3} + \frac{3\mu}{\|\bar{S}_k\|^5}\bar{S}_k\bar{S}_k^{\mathrm{T}} - \Omega^2\right) & -2\Omega \end{bmatrix}. \tag{47}$$

The performances of ESKF and traditional SVF are compared in Figures 2–4. According to Figure 2, both ESKF and traditional SVF perform well when the size of uncertainty is small (C1). We increase the size of the uncertainty from C1 to C2 with the fixed parameters of the filters. Figure 3 shows that the estimation error of ESKF remains small while the estimation performance of traditional SVF deteriorates. We continue to increase the size of the uncertainty to an unbounded uncertainty as C3. Figures 4 and 5 show that although unboundedness of $\Delta_F(t)$ leads to the divergence of system, the estimation of ESKF performs well while the estimate of traditional SVF becomes divergent. It indicates that ESKF achieves better filtering results than traditional SVF in the case of large uncertainty or even unbounded uncertainty. Furthermore, the uncertainty F_k can be timely estimated by ESKF for improving the estimation performance, as the estimation performance of F_k shown in Figures 2–4. This is the reason why ESKF can tolerate larger uncertainties than the traditional SVF. Additionally, Figure 6 presents the curves of the diagonal elements of P_k and the curves of the mean square estimation errors (MSE) for ESKF obtained from simulation for 100 times in C2. Obviously, the MSE of ESKF, being accordance with our theoretical result, is bounded by the diagonal elements of P_k for ESKF. Besides, as the curves of P_k elements can better reflect the MSE of ESKF, P_k can be treated as the accuracy evaluation of ESKF.

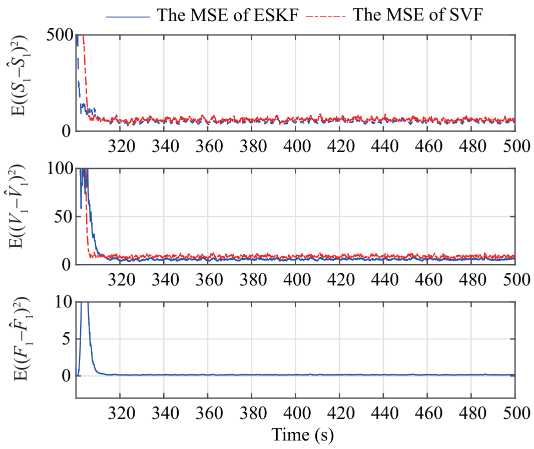

Figure 2 (Color online) The MSE of ESKF and SVF for C1 (Example 2).

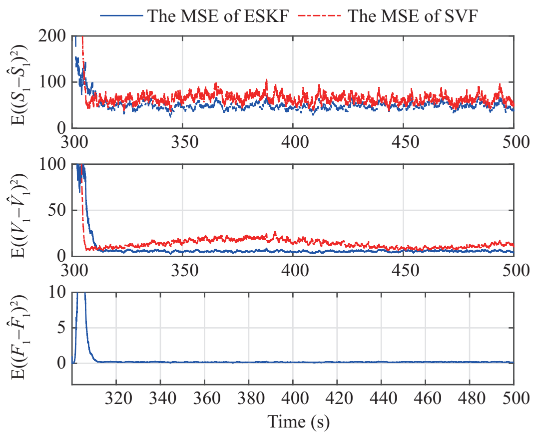

Figure 3 (Color online) The estimate error of ESKF and SVF for C2 (Example 2).

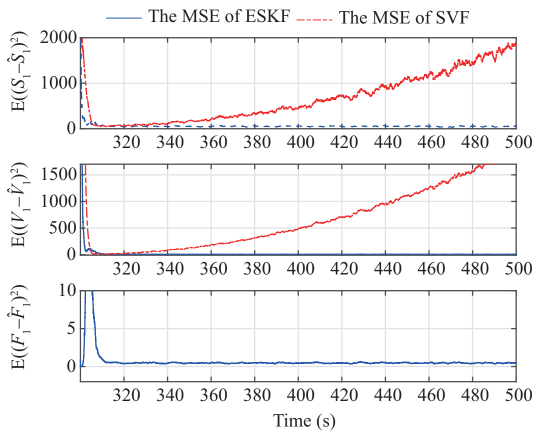

Figure 4 (Color online) The MSE of ESKF and SVF for C3 (Example 2).

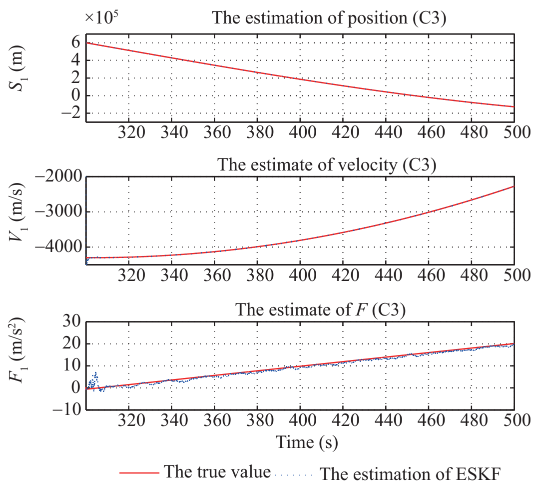

Figure 5 (Color online) The estimation of ESKF and the state value for C3 (Example 2).

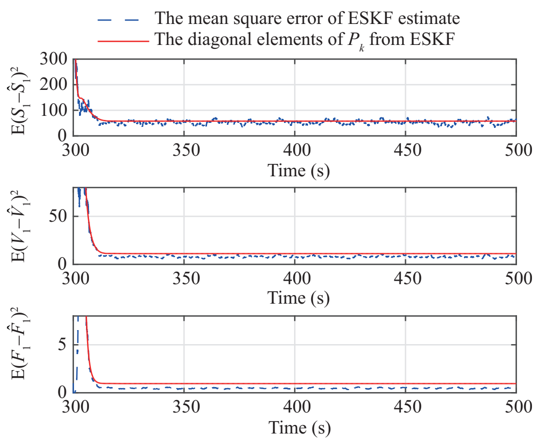

Figure 6 (Color online) The mean square error of ESKF and the diagonal elements of P_k in ESKF for C3 (Example 2).

5 Conclusion

This paper proposes the ESKF aiming to solve the filtering problem for a class of nonlinear time-varying uncertain systems. The essence of ESKF is to actively estimate the "total disturbance", which lumps model uncertainties and unknown disturbances, for better robustness and higher estimation precision. Thus the stability of ESKF can be assured for uncertain nonlinear systems under a weaker condition, rather than the conditions of true model, small enough initial estimation error and noise. Additionally, the paper shows that the estimation precision can be timely evaluated through the filter parameters. Moreover, when the uncertainty is constant, ESKF asymptotically tends to the unbiased minimum variance filter. The simulation results on two typical examples illustrate the effectiveness of the proposed filter.

This paper offers a new idea to handle nonlinearities and uncertainties by the extended state design of filters. Besides, the paper also optimizes the gains of ESO for stochastic uncertain system. We believe this work will have a huge impact on both nonlinear filter and ESO.

Acknowledgements This work was supported by National Natural Science Foundation of China (NSFC) (Grant Nos. 61603380, 61633003-3), and National Basic Research Program of China (Grant No. 2014CB845301). The authors would like to thank professer Lei GUO with Chinese Academy of Sciences for his helpful suggestions.

References

1 Anderson B D O, Moore J B. Optimal Filtering. Upper Saddel River: Prentice Hall, 1995
2 Grewal M S, Andrews A P. Kalman Filtering: Theorey and Practice. Upper Saddel River: Prentice Hall, 1993
3 Guo L. Estimating time-varying parameters by the Kalman filter based algorithm: stability and convergence. IEEE Trans Autom Control, 1990, 35: 141–147
4 Chen C, Liu Z X, Guo L. Performance bounds of distributed adaptive filters with cooperative correlated signals. Sci China Inf Sci, 2016, 59: 112202
5 Ljung L. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. IEEE Trans Autom Control, 1979, 24: 36–50
6 Deyst J, Price C. Conditions for asymptotic stability of the discrete minimum-variance linear estimator. IEEE Trans Autom Control, 1968, 13: 702–705
7 Reif K, Gunther S, Yaz E, et al. Stochastic stability of the discrete-time extended Kalman filter. IEEE Trans Autom Control, 1999, 44: 714–728
8 Reif K, Gunther S, Yaz E, et al. Stochastic stability of the continuous-time extended Kalman filter. IEEE Proc Control Theory Appl, 2000, 147: 45–52
9 Julier S J, Uhlmann J K. Unscented filtering and nonlinear estimation. Proc IEEE, 2004, 92: 401–422
10 Zhang W, Chen B S, Tseng C S. Robust H_∞ filtering for nonlinear stochastic systems. IEEE Trans Signal Process, 2005, 53: 589–598
11 Zhu Y M. From Kalman filtering to set-valued filtering for dynamic systems with uncertainty. Commun Inf Syst, 2012, 12: 97–130
12 Zorzi M. Robust Kalman filtering under model perturbations. IEEE Trans Autom Control, 2017, 62: 2902–2907
13 Xie L, Soh Y C, de Souza C E. Robust Kalman filtering for uncertain discrete-time systems. IEEE Trans Autom Control, 1994, 39: 1310–1314
14 Liu Q Y, Wang Z D, He X, et al. A resilient approach to distributed filter design for time-varying systems under stochastic nonlinearities and sensor degradation. IEEE Trans Signal Process, 2017, 65: 1300–1309
15 Simon D. Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Hoboken: John Wiley & Sons, 2006
16 Shaked U, Berman N. H_∞ nonlinear filtering of discrete-time processes. IEEE Trans Signal Process, 1995, 43: 2205–2209
17 Su Q Y, Huang Y, Jiang Y G, et al. Quasi-consistent fusion navigation algorithm for DSS. Sci China Inf Sci, 2018, 61: 012201
18 Sayed A H. A framework for state-space estimation with uncertain models. IEEE Trans Autom Control, 2001, 46: 998–1013
19 Seo J, Yu M J, Park C G, et al. An extended robust H_∞ filter for nonlinear constrained uncertain systems. IEEE Trans Signal Process, 2006, 54: 4471–4475
20 Scholte E, Campbell M E. A nonlinear set membership filter for online applications. Int J Robust Nonlin Control, 2003, 13: 1337–1358
21 Huang Y, Xu K K, Han J Q, et al. Flight control design using extended state observer and non-smooth feedback. In: Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, 2001. 223–228

22 Liu L, Wang D, Peng Z H, et al. Saturated coordinated control of multiple underactuated unmanned surface vehicles over a closed curve. Sci China Inf Sci, 2017, 60: 070203
23 Tang S, Zhang L, Qian S K, et al. Second-order sliding mode attitude controller design of a small-scale helicopter. Sci China Inf Sci, 2016, 59: 112209
24 Su J B, Ma H Y, Qiu W B, et al. Task-independent robotic uncalibrated hand-eye coordination based on the extended state observer. IEEE Trans Syst Man Cybern, 2004, 34: 1917–1922
25 Liu H X, Li S H. Speed control for PMSM servo system using predictive functional control and extended state observer. IEEE Trans Ind Electron, 2012, 59: 1171–1183
26 Dong L, Avanesian D. Drive-mode control for vibrational MEMS gyroscopes. IEEE Trans Ind Electron, 2009, 56: 956–963
27 Sun L, Dong J Y, Li D H, et al. A practical multivariable control approach based on inverted decoupling and decentralized active disturbance rejection control. Ind Eng Chem Res, 2016, 55: 2008–2019
28 Gao Z Q. Scaling and bandwidth-parameterization based controller tuning. In: Proceedings of the American Control Conference, Denver, 2003
29 Bai W Y, Xue W C, Huang Y, et al. The extended state filter for a class of multi-input multi-output nonlinear uncertain hybrid systems. In: Proceedings of the 33rd Chinese Control Conference (CCC), Nanjing, 2014. 2608–2613
30 Bai W Y, Xue W C, Huang Y, et al. Extended state filter design for general nonlinear uncertain systems. In: Proceedings of the 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Hangzhou, 2015. 712–717
31 Guo B Z, Zhao Z L. On the convergence of an extended state observer for nonlinear systems with uncertainty. Syst Control Lett, 2011, 60: 420–430
32 Han J Q. From PID to active disturbance rejection control. IEEE Trans Ind Electron, 2009, 56: 900–906
33 Bougerol P. Kalman filtering with random coefficients and contractions. SIAM J Control Optim, 1993, 31: 942–959
34 Zhao C, Guo L. PID controller design for second order nonlinear uncertain systems. Sci China Inf Sci, 2017, 60: 022201
35 Bar-Shalom Y, Li X R, Kirubarajan T. Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. Hoboken: John Wiley & Sons, 2004
36 Shamma J S, Tu K Y. Approximate set-valued observers for nonlinear systems. IEEE Trans Autom Control, 1997, 42: 648–658
37 Zhou B, Qian K, Ma X D, et al. A new nonlinear set membership filter based on guaranteed bounding ellipsoid algorithm. Acta Autom Sin, 2013, 39: 146–154
38 Li X R, Jilkov V P. A survey of maneuvering target tracking — part II: ballistic target models. In: Proceedings of SPIE Conference on Signal and Data Processing of Small Targets, San Diego, 2001
39 Austin J W, Leondes C T. Statistically linearized estimation of reentry trajectories. IEEE Trans Aerosp Electron Syst, 1981, 54–61
40 Costa P J. Adaptive model architecture and extended Kalman-Bucy filters. IEEE Trans Aerosp Electron Syst, 1994, 30: 525–533

Appendix A Proof of Lemma 1

Sufficiency. If (A_k, C_k) is uniformly observable, then for any $\lambda \in \mathbb{R}$, $\mathrm{rank}\begin{bmatrix}\lambda I - A_k \\ C_k\end{bmatrix} = n + l,\ \forall k = 0, 1, \ldots$.

From (3), there is

$$\mathrm{rank}\begin{bmatrix}\lambda I - \bar{A}_k & -\bar{B}_k \\ 0 & \lambda I - I \\ \bar{C}_k & 0\end{bmatrix} = n + l,\ \forall \lambda \in \mathbb{R},\ \forall k = 0, 1, \ldots. \tag{A1}$$

Thus

$$\mathrm{rank}\begin{bmatrix}\lambda I - \bar{A}_k \\ 0 \\ \bar{C}_k\end{bmatrix} = n,\ \forall \lambda \in \mathbb{R},\ \forall k = 0, 1, \ldots. \tag{A2}$$

Therefore, $(\bar{A}_k, \bar{C}_k)$ is uniformly observable.

In addition, let $\lambda = 1$, there is

$$\mathrm{rank}\begin{bmatrix}I - \bar{A}_k & -\bar{B}_k \\ 0 & 0 \\ \bar{C}_k & 0\end{bmatrix} = n + l,\ \forall k = 0, 1, \ldots. \tag{A3}$$

Necessity. Otherwise, if $(\bar{A}_k, \bar{C}_k)$ is uniformly observable and $\mathrm{rank}\begin{bmatrix}I - \bar{A}_k & -\bar{B}_k \\ \bar{C}_k & 0\end{bmatrix} = n + l$, for any $\lambda \in \mathbb{R}$, there is

$$\mathrm{rank}\begin{bmatrix}\lambda I - \bar{A}_k \\ \bar{C}_k\end{bmatrix} = n,\ \forall k = 0, 1, \ldots. \tag{A4}$$

Thus for any $\lambda \in \mathbb{R}$ and $\lambda \neq 1$, there is

$$\operatorname{rank}\begin{bmatrix} \lambda I - \bar{A}_k & -\bar{B}_k \\ 0 & \lambda I - I \\ \bar{C}_k & 0 \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \lambda I - \bar{A}_k & 0 \\ 0 & \lambda I - I \\ \bar{C}_k & 0 \end{bmatrix} = n + l,\ \forall k = 0, 1, \ldots. \tag{A5}$$

Besides, if $\lambda = 1$, there is

$$\operatorname{rank}\begin{bmatrix} \lambda I - \bar{A}_k & -\bar{B}_k \\ 0 & \lambda I - I \\ \bar{C}_k & 0 \end{bmatrix} = \operatorname{rank}\begin{bmatrix} I - \bar{A}_k & -\bar{B}_k \\ 0 & 0 \\ \bar{C}_k & 0 \end{bmatrix} = n + l. \tag{A6}$$

In conclusion, for any $\lambda \in \mathbb{R}$, we have

$$\operatorname{rank}\begin{bmatrix} \lambda I - \bar{A}_k & -\bar{B}_k \\ 0 & \lambda I - I \\ \bar{C}_k & 0 \end{bmatrix} = n + l,\ \forall k = 0, 1, \ldots. \tag{A7}$$

Therefore (A_k, C_k) is uniformly observable.

Appendix B Proof of Theorem 1

From (19), it is obvious that the selection methods of the parameters $P_0,\ R_k,\ Q_{1,k},\ Q_{2,k},\ \theta$ endow ESKF with the consistence property, that is, $E(e_k e_k^{\mathrm{T}}) \leqslant P_k$.

Next, we will prove that $\{P_k\}_{k=1}^{\infty}$ is uniformly bounded.

From (27) P_k can be expressed as

$$\begin{aligned} P_{k+1} = &\left(\sqrt{1+\theta}A_k + K_k\sqrt{1+\theta}C_k\right) P_k \left(\sqrt{1+\theta}A_k + K_k\sqrt{1+\theta}C_k\right)^{\mathrm{T}} \\ &+ K_k R_k K_k{}^{\mathrm{T}} + \left(1 + \frac{1}{\theta}\right) Q_{1,k} + Q_{2,k}. \end{aligned} \tag{B1}$$

Since $(\sqrt{1+\theta}A_k, \sqrt{1+\theta}C_k)$ is uniformly observable and $\{R_k\}_{k=1}^{\infty},\ \{Q_{1,k}\}_{k=1}^{\infty},\ \{Q_{2,k}\}_{k=1}^{\infty}$ are uniformly bounded, $\{P_k\}_{k=1}^{\infty}$ is uniformly bounded[3)].

Appendix C Proof of Lemma 2

Assume P_1^* and P_1^{**} denote the P_1 under the situations of $(P_0^*, Q_{1,0}^*, Q_{2,0}^*, R_0^*)$ and $(P_0^{**}, Q_{1,0}^{**}, Q_{2,0}^{**}, R_0^{**})$ in (25)–(28), respectively, and the corresponding K are K^* and K^{**} respectively. From

$$P_0{}^* < P_0{}^{**}, \quad Q_{1,0}^* < Q_{1,0}^{**}, \quad Q_{2,0}^* < Q_{2,0}^{**}, \quad R_0^* < R_0^{**},$$

it is obvious that

$$\begin{aligned} P_1^{**} = &(1+\theta)\left(A_0 + K_0^{**}C_0\right) P_0^{**} \left(A_0 + K_0^{**}C_0\right)^{\mathrm{T}} + K_0^{**} R_0^{**} K_0^{**\mathrm{T}} + \left(1 + \frac{1}{\theta}\right) Q_{1,0}^{**} + Q_{2,0}^{**} \\ > &(1+\theta)\left(A_0 + K_0^{**}C_0\right) P_0^{*} \left(A_0 + K_0^{**}C_0\right)^{\mathrm{T}} + K_0^{**} R_0^{*} K_0^{**\mathrm{T}} + \left(1 + \frac{1}{\theta}\right) Q_{1,0}^{*} + Q_{2,0}^{*}. \end{aligned} \tag{C1}$$

From (23), we can get

$$K_0^* = \arg\min_{K_0} \left\{ (1+\theta)\left(A_0 + K_0 C_0\right) P_0^* \left(A_0 + K_0 C_0\right)^{\mathrm{T}} + K_0 R_0^* K_0{}^{\mathrm{T}} + \left(1 + \frac{1}{\theta}\right) Q_{1,0}^* + Q_{2,0}^* \right\}. \tag{C2}$$

Thus

$$P_1^{**} > (1+\theta)\left(A + K_0^* C\right) P_0^* \left(A + K_0^* C\right)^{\mathrm{T}} + K_0^* R_0^* K_0^{*\mathrm{T}} + \left(1 + \frac{1}{\theta}\right) Q_{1,0}^* + Q_{2,0}^* = P_1^*. \tag{C3}$$

Therefore, $P_1^* < P_1^{**}$. Repeating this procedure, we get

$$P_k^* < P_k^{**}. \tag{C4}$$

3) Anderson B D O, Moore J B. Detectability and stabilizability of time-varying discrete-time linear systems. SIAM J Control Optim, 1981, 19: 20–32.

Appendix D Proof of Theorem 2

For the constant $\bar{A},\ \bar{B},\ \bar{C},\ R$, the unbiased minimun variance estimation (the estimation of KF) for $[X_k^{\rm T}\ \ F_k^{\rm T}]^{\rm T}$ satisfies

$$\begin{bmatrix}\bar{X}_{k+1}\\ \bar{F}_{k+1}\end{bmatrix} = A\begin{bmatrix}\bar{X}_k\\ \bar{F}_k\end{bmatrix} + \bar{K}_k\left(Y_k - C\begin{bmatrix}\bar{X}_k\\ \bar{F}_k\end{bmatrix}\right), \tag{D1}$$

$$\bar{K}_k = -A\bar{P}_kC^{\rm T}(C\bar{P}_kC^{\rm T}+R)^{-1}, \tag{D2}$$

$$\bar{P}_{k+1} = (A+\bar{K}_kC)\bar{P}_k(A+\bar{K}_kC)^{\rm T} + \bar{K}_kR\bar{K}_k^{\rm T} + Q_2, \tag{D3}$$

$$\begin{bmatrix}\bar{X}_0\\ \bar{F}_0\end{bmatrix} = \begin{bmatrix}E(X_0)\\ F_0\end{bmatrix},\quad \bar{P}_0 = \begin{bmatrix}E\left((X_0-\bar{X}_0)(X_0-\bar{X}_0)^{\rm T}\right) & 0\\ 0 & 0\end{bmatrix}. \tag{D4}$$

According to the property of KF [1], $\overline{P}_k$ is the minimun variance of $[X_k^{\rm T}\ \ F_k^{\rm T}]^{\rm T}$. From calculating, we can get the limit of $\bar{P}_k$ satisfying

$$\lim_{k\to\infty}\bar{P}_k = \bar{P} = \begin{bmatrix}\bar{P}_{(1)} & 0\\ 0 & 0\end{bmatrix}, \tag{D5}$$

where $\bar{P}_{(1)}$ is the unique positive solution for the following Riccati equation:

$$\bar{P}_{(1)} = -\bar{A}\bar{P}_{(1)}\bar{C}^{\rm T}(\bar{C}\bar{P}_{(1)}\bar{C}^{\rm T}+R)^{-1}\bar{C}\bar{P}_{(1)}\bar{A}^{\rm T} + \bar{A}\bar{P}_{(1)}\bar{A}^{\rm T} + S, \tag{D6}$$

provided that $(\bar{A},\bar{C},S)$ satisfies the observability and controllability conditions [1].

As for ESKF, from $F(X(t),t)\equiv F_0$ it can be verified that

$$G_k = 0,\quad \overline{G}(X_k,k) = 0,\quad \forall k>0. \tag{D7}$$

Thus ESKF with $\overline{Q}_k = 0$ becomes

$$\begin{bmatrix}\hat{X}_{k+1}\\ \hat{F}_{k+1}\end{bmatrix} = A\begin{bmatrix}\hat{X}_k\\ \hat{F}_k\end{bmatrix} - K_k\left(Y_k - C\begin{bmatrix}\hat{X}_k\\ \hat{F}_k\end{bmatrix}\right), \tag{D8}$$

$$K_k = -AP_kC^{\rm T}(CP_kC^{\rm T}+R)^{-1}, \tag{D9}$$

$$P_{k+1} = (A+K_kC)\,P_k\,(A+K_kC)^{\rm T} + K_kRK_k{}^{\rm T} + Q_2, \tag{D10}$$

which equals to KF algorithm. However, since F_0 is unknown, the initial estimation $\hat{F}_0$ of ESKF may be biased. In this case, we will prove that ESKF (D8) tends to the unbiased minimum variance estimation, i.e.,

$$\lim_{k\to\infty}E(e_ke_k^{\rm T}) = \bar{P},\quad \lim_{k\to\infty}E(e_k) = 0. \tag{D11}$$

Since P_k is bounded (see Theorem 1), there exists a converged subsequence $\{P_{k_j}\}_{j=1}^{\infty}$ such that

$$\lim_{j\to\infty}P_{k_j} = P. \tag{D12}$$

Besides, P satisfies the following equation:

$$P = -APC^{\rm T}(CPC^{\rm T}+R)^{-1}CPA^{\rm T} + APA^{\rm T} + Q_2. \tag{D13}$$

According to the expression of $A,\ C,\ Q_2$, we introduce the natural block stucture

$$P = \begin{bmatrix}P_{(1,1)_{n\times n}} & P_{(1,2)_{n\times l}}\\ P^{\rm T}_{(1,2)_{l\times n}} & P_{(2,2)_{l\times l}}\end{bmatrix}. \tag{D14}$$

Thus, (D13) can be rewritten as

$$\begin{aligned}P_{(1,1)} =&\bar{A}P_{(1,1)}\bar{A}^{\rm T} + \bar{A}P_{(1,2)}\bar{B}^{\rm T} + \bar{B}P_{(1,2)}^{\rm T}\bar{A}^{\rm T} + \bar{B}P_{(2,2)}\bar{B}^{\rm T} + S\\ &-(\bar{A}P_{(1,1)}\bar{C}^{\rm T} + \bar{B}P_{(1,2)}^{\rm T}\bar{C}^{\rm T})(\bar{C}P_{(1,1)}\bar{C}^{\rm T}+R)^{-1}(\bar{A}P_{(1,1)}\bar{C}^{\rm T} + \bar{B}P_{(1,2)}^{\rm T}\bar{C}^{\rm T})^{\rm T},\end{aligned} \tag{D15}$$

$$P_{(1,2)} = -(\bar{A}P_{(1,1)}\bar{C}^{\rm T} + \bar{B}P_{(1,2)}^{\rm T}\bar{C}^{\rm T})(\bar{C}P_{(1,1)}\bar{C}^{\rm T}+R)^{-1}\bar{C}P_{(1,2)} + \bar{A}P_{(1,2)} + \bar{B}P_{(2,2)}, \tag{D16}$$

$$P_{(2,2)} = P_{(2,2)} - P_{(1,2)}^{\rm T}\bar{C}^{\rm T}(\bar{C}P_{(1,1)}\bar{C}^{\rm T}+R)^{-1}\bar{C}P_{(1,2)}. \tag{D17}$$

From (D17) and $\bar{C}P_{(1,1)}\bar{C}^{\rm T}+R>0$, we known that

$$\bar{C}P_{(1,2)} = 0. \tag{D18}$$

Then substitute (D18) into (D16), we have

$$(I-\bar{A})P_{(1,2)} - \bar{B}P_{(2,2)} = 0. \tag{D19}$$

Since (A, C) is observable, for any $\lambda \in \mathbb{R}$, $\begin{bmatrix} \lambda I - A \\ C \end{bmatrix}$ is full column rank. Take $\lambda = 1$ so that $\begin{bmatrix} I - \bar{A} & -\bar{B} \\ \bar{C} & 0 \end{bmatrix}$ is full column rank. Therefore, we can conclude from (D18)–(D19) that

$$P_{(1,2)} = 0, \quad P_{(2,2)} = 0. \tag{D20}$$

Substituting (D20) into (D15), we can get $P_{(1,1)}$ satisfying

$$P_{(1,1)} = -\bar{A}P_{(1,1)}\bar{C}^{\mathrm{T}}(\bar{C}P_{(1,1)}\bar{C}^{\mathrm{T}} + R)^{-1}\bar{C}P_{(1,1)}\bar{A}^{\mathrm{T}} + \bar{A}_k P_{(1,1)}\bar{A}^{\mathrm{T}} + S. \tag{D21}$$

Since $(\bar{A}, \bar{C}, S)$ satisfies the observability and controllability conditions, $P_{(1,1)} = \bar{P}_{(1)}$ is the unique positive solution for (D21). Hence,

$$P = \bar{P}. \tag{D22}$$

To conclude, all the converged subsequences of $\{P_k\}_{k=0}^{\infty}$ converge to $\bar{P}$. Next, we will prove

$$\lim_{k\to\infty} P_k = \bar{P}. \tag{D23}$$

Let $\bigcup_{i=1}\{P_{k_j}^{(i)}\}_{j=0}^{\infty}$ denote the union of the converged subsequences of $\{P_k\}_{k=0}^{\infty}$. Suppose

$$\{P_k\}_{k=0}^{\infty} \backslash \left\{ \bigcup_{i=1} \left\{ P_{k_j}^{(i)} \right\}_{j=0}^{\infty} \right\} \neq \emptyset.$$

Since $\{P_k\}_{k=0}^{\infty} \backslash \{\bigcup_{i=1}\{P_{k_j}^{(i)}\}_{j=0}^{\infty}\}$ is consistent bounded, there exists a subsequence satisfying

$$\{P_{(k_i)}^{*}\}_{i=1}^{\infty} \subseteq \{P_k\}_{k=0}^{\infty} \backslash \left\{ \bigcup_{i=1} \left\{ P_{k_j}^{(i)} \right\}_{j=0}^{\infty} \right\},$$

such that

$$\lim_{i\to\infty} P_{(k_i)}^{*} = \bar{P}.$$

This contradicts $P_{(k_i)}^{*} \in \{P_k\}_{k=0}^{\infty} \backslash \{\bigcup_{i=1}\{P_{k_j}^{(i)}\}_{j=0}^{\infty}\}$. Thus

$$\{P_k\}_{k=0}^{\infty} \backslash \left\{ \bigcup_{i=1} \left\{ P_{k_j}^{(i)} \right\}_{j=0}^{\infty} \right\} = \emptyset.$$

That is, every subsequence of $\{P_k\}_{k=0}^{\infty}$ converges to $\bar{P}$, which yields (D23). From the property of $\overline{P}_k$ and P_k, the error covariance of ESKF satisfies

$$\overline{P}_k \leqslant E(e_k e_k^{\mathrm{T}}) \leqslant P_k. \tag{D24}$$

Therefore, from (D5) and (D23),

$$\lim_{k\to\infty} E(e_k e_k^{\mathrm{T}}) = \bar{P}. \tag{D25}$$

Since KF (D1)–(D4) is an unbiased minimum variance estimation with the covariance being $\overline{P}_k$, it is easy to see that the estimation $[X_k^{\mathrm{T}} \;\; F_k^{\mathrm{T}}]^{\mathrm{T}}$ is asymptotic unbiased, that is

$$\lim_{k\to\infty} E(e_k) = 0. \tag{D26}$$