

## UKF Based Nonlinear Filtering for Parameter Estimation in Linear Systems with Correlated Noise

Jiahe Xu\*, Tatjana Kolemisevska-Gugulovska\*\*, Xiuping Zheng\*, Yuanwei Jing\*,  
Georgi M. Dimirovski\*\*\*, Member, IEEE

\* Institute of Information Science and Engineering,  
Northeastern University, Shenyang, Liaoning, 110004, P.R. of China  
(e-mail: ellipsis@163.com; ywjing@mail.neu.edu.cn)

\*\* Faculty of EE&IT, SS Cyril and Methodius University, 1000 Skopje, Rep. of Macedonia  
(e-mail: tanjakg@feit.ukim.edu.mk)

\*\*\* Dogus University of Istanbul, Faculty of Engineering, TR-347222 Istanbul,  
Rep. of Turkey, and with SS (e-mail: gdimirovski@dogus.edu.tr)

---

**Abstract:** Based on the Unscented Kalman Filter (UKF), the nonlinear filter is presented for parameter estimation in linear system with correlated noise where the unknown parameters are estimated as a part of an enlarged state vector. To avoid the computational burden in determining the state estimates when only the parameter estimates are required, a new form of UKF, where the state consists only of the parameters to be estimated, is proposed. The algorithm is based on the inclusion of the computed residuals in the observation matrix of a state representation of the system. Convergence properties of the proposed algorithm are analyzed and ensured. The algorithm is verified by using Matlab simulations on the vehicle navigation systems with aided GPS.

---

### 1. INTRODUCTION

The problem of parameter estimation in stochastic linear dynamic systems has been researched considerable because of its importance in model building and control theory (Astrom *et al.*, 1971; Granados *et al.*, 1998; James *et al.*, 2000; Chin-Lang Tsai *et al.*, 2006). It is well known that the general case of correlated noise leads to a nonlinear estimation problem. A number of methods have been proposed, most of which base on the Extended Kalman Filter (EKF) (Jazwinski, 1970; Astrom *et al.*, 1971). A variant of method will be discussed in this paper.

The approach to parameter estimation in linear systems can be summarized as follows. To estimate the unknown parameters of a linear system, the parameters are appended to the state variables, and a state estimator for the enlarged system is then used to obtain joint estimates of both the original system state and the system parameters. However, since the enlarged system is nonlinear (Astrom *et al.*, 1971), to make the computation of the estimates feasible, the unscented kalman filter (UKF) (Julier *et al.*, 1995), which aims at the nonlinear system directly (Julier, 2000; Julier *et al.*, 2000; Lefebvre *et al.*, 2002; Julier *et al.*, 2004), is proposed.

Although the UKF approach to parameter estimation has a strong intuitive appeal and offers the possibility (Julier *et al.*, 1997a; Julier *et al.*, 1997b; Julier *et al.*, 1997c; Wan *et al.*, 2000), it suffers from many disadvantages (Pan Quan *et al.*, 2005). The computational burden of estimating the enlarged state and magnified by the necessity for Unscented Transform (UT) at each step may have divergence problems, and although filter divergence can be avoided in modified

algorithms, the modifications require additional computing time.

The contribution of this paper rests in the development of the UKF for parameter estimation. This is achieved by modeling the system, whose parameters are to be estimated, by state equations where the state consists of the system and noise parameters, while the corresponding inputs, outputs, and computed residuals are collected in the observation matrix of the state equations. This results again in a nonlinear system, and the UKF state estimator directly gives, in this case, the parameter estimates only. To study the convergence properties of the algorithm, some techniques based on the associated differential equation (Ljung, 1977; Ljung, 1979; Xiong *et al.*, 2006) are used. It is shown that the UKF can achieve the convergence properties.

The paper is structured as follows. Section II contains the outline of the problem and gives the UKF equations. In Section III, the parameter estimation algorithm is developed, and the related convergence results are obtained in Section IV. The algorithm is verified by using Matlab simulations on the vehicle navigation systems with aided GPS (Nebot, 1998) and the results show the effectiveness of the algorithm.

### 2. PRELIMINARIES

Consider the state representation of a multivariable system

$$\begin{aligned}x(t+1) &= A_p x(t) + B_p u(t) + v(t) \\ y(t) &= C_p x(t) + w(t)\end{aligned}\quad (1)$$

where  $u(t)$ ,  $y(t)$ , and  $x(t)$  are the input, output, and state vector of dimensions  $n_u$ ,  $n_y$ , and  $n_x$ , respectively, and  $\{v(t)\}$ ,  $\{w(t)\}$  are sequences of independent random vectors with zero mean and the covariance.

$$E\left[\begin{matrix} w(i) \\ v(i) \end{matrix}\right] \begin{bmatrix} w^T(j) & v^T(j) \end{bmatrix} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{ij}, \quad \delta_{ii} = 1, \quad \delta_{ij} = 0 \quad (i \neq j) \quad (2)$$

Also, the initial state  $x(0)$ , assumed to be a random vector with zero mean and a covariance  $\Pi(0)$  is considered independent of  $\{v(t)\}$ ,  $\{w(t)\}$  for  $t > 0$ , and the matrices in (1) and (2) are time invariant.

We shall now consider the case where the matrices in (1) and (2) are unknown and to be estimated from the input-output measurements  $\{u(t), y(t)\}$ ,  $t = 0, 1, \dots$ . For this purpose, we shall establish a model of (1):

$$\begin{aligned} x(t+1) &= A(\theta)x(t) + B(\theta)u(t) + v(t) \\ y(t) &= C(\theta)x(t) + w(t) \end{aligned} \quad (3)$$

where

$$\begin{aligned} E\left\{\begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w^T(j) & v^T(j) \end{bmatrix}\right\} &= \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{ij}, \delta_{ii} = 1, \delta_{ij} = 0 (i \neq j) \\ E(x(0)) &= 0, \quad E(x(0)x^T(0)) = \Pi(0) \end{aligned} \quad (4)$$

In this model, the unknown matrices are represented as some differentiable functions of a parameter vector  $\theta$  which will be estimated by using the UKF approach. First, form an enlarged state vector by appending the parameter vector  $\theta = \theta(t)$  to the state

$$z(t) = [x(t) \quad \theta(t)]^T \quad (5)$$

The resulting state equations are nonlinear:

$$\begin{aligned} z(t+1) &= f(z(t), u(t)) + \begin{bmatrix} v(t) \\ 0 \end{bmatrix} \\ y^2(t) &= g(z(t)) + w(t) \end{aligned} \quad (6)$$

where

$$f(z(t), u(t)) = \begin{bmatrix} A(\theta)x(t) + B(\theta)u(t) \\ \theta(t) \end{bmatrix} \quad (7)$$

$$g(z(t)) = C(\theta)x(t) \quad (8)$$

The required parameter estimates are now obtained by applying the UKF to (6) which gives.

The  $n$ -dimensional random variable  $z(t)$  with mean  $\hat{z}(t)$  and covariance  $P^z(t)$  can be approximated by sigma points  $\xi_i(t)$  selected from the columns of  $\hat{z}(t) \pm (a\sqrt{LP^z(t)})_i$ ,  $i = 0, \dots, 2L$ . The opposite weight  $\omega_i$  is  $\omega_i(0) = 1 - (1/a^2)$ ,  $\omega_i(i) = 1/2La^2$  ( $i = 1, 2, \dots, 2L$ ).

The predicted mean and covariance are computed as

$$\hat{\xi}_i(t+1|t) = f(\xi_i(t), u(t)), \quad \hat{z}(t+1|t) = \sum_{i=0}^{2n} \omega_z(i) \xi_i(t+1|t)$$

$$P^z(t+1|t) = \sum_{i=0}^{2n} \omega_z(i) (\xi_i(t+1|t) - \hat{z}(t+1|t)) (\xi_i(t+1|t) - \hat{z}(t+1|t))^T + Q'$$

$$Q' = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$$

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

$$y_i^z(t+1|t) = g[\xi_i(t+1|t)], \quad \hat{y}^z(t+1|t) = \sum_{i=0}^{2n} \omega_z(i) y_i^z(t+1|t)$$

$$P_{yy}^z(t+1|t) = \sum_{i=0}^{2n} \omega_z(i) (y_i^z(t+1|t) - \hat{y}^z(t+1|t)) (y_i^z(t+1|t) - \hat{y}^z(t+1|t))^T + R$$

$$P_{zy}^z(t+1|t) = \sum_{i=0}^{2n} \omega_z(i) (\xi_i(t+1|t) - \hat{z}(t+1|t)) (y_i^z(t+1|t) - \hat{y}^z(t+1|t))^T$$

$$K^z(t+1) = P_{zy}^z(t+1|t) [P_{yy}^z(t+1|t)]^{-1}$$

$$\hat{z}(t+1) = \hat{z}(t+1|t) + K^z(t+1) (y^z(t+1) - \hat{y}^z(t+1|t)), \quad \hat{z}(0) = \hat{z}^0 = [0 \quad \hat{\theta}^0]^T$$

$$P^z(t+1) = P^z(t+1|t) - K^z(t+1) P_{yy}^z(t+1|t) [K^z(t+1)]^T,$$

$$P(0) = P^0 = \begin{bmatrix} \Pi(\hat{\theta}^0) & 0 \\ 0 & \Sigma^0 \end{bmatrix}$$

with  $\hat{\theta}^0$ ,  $\Sigma^0$  incorporating any a priori information we may have about the parameter vector.

When the algorithm are implemented, a fairly complex algorithm results, and for high-order systems, the computational burden may be substantial. Even worse, the usual arbitrary choice of the model covariance matrices (4)

may lead to divergence and bias problems. Therefore, it would be of interest to investigate an alternative representation to (1) which should not give rise to the need for the selection of the covariance matrices (4) when there is no a priori information about their structure available; the representation should have fewer parameters and should obviate the necessity for the state estimation where only the parameter estimates are required. At the same time, we would like to keep the structure of the Kalman filter estimator because of its advantageous minimum variance formulation of the estimation problem. A representation of (1) having all the above features will be discussed in the next section and will form a basis for the proposed estimation algorithm.

### 3. THE ALGORITHM

Since in parameter estimation only the input-output properties of a system are of interest, we shall now start with the parameter parsimonious vector difference equation (VDE) representation of a multivariable linear system. We shall use the form

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) \quad (8)$$

where  $q^{-1}$  is the backward shift operator.

$$q^{-1}y(t) = y(t-1)$$

and

$$\begin{aligned} A(q^{-1}) &= I + A_1 q^{-1} + \dots + A_{n_a} q^{-n_a} \\ B(q^{-1}) &= B_1 q^{-1} + \dots + B_{n_b} q^{-n_b} \\ C(q^{-1}) &= I + C_1 q^{-1} + \dots + C_{n_c} q^{-n_c} \end{aligned} \quad (9)$$

where the coefficients  $A_i$ ,  $C_i$  are  $n_y \times n_y$  matrices, whereas the coefficients  $B_i$  are  $n_y \times n_u$  matrices. The sequence  $\{e(t)\}$  is assumed to consist of independent random vectors of dimension  $n_y$ , having zero mean and a covariance

$$E\{e(t)e^T(l)\} = \Lambda_0 \delta_{tl} \quad (10)$$

Although it is possible to obtain under some assumptions the representation (8) from the state form (1) and vice versa, in the sequel we shall assume that the data  $\{u(t), y(t)\}$ ,  $t = 0, 1, \dots$  have been generated by (8), which is an important form in its own right, and we shall formulate the problem as the estimation of the parameters in (9) and (10) from the given input-output data.

To develop the estimation algorithm, first write the system (8) in the form

$$y^\theta(t) = \psi_\theta(t) \theta_0 + e(t) \quad (11)$$

with

$$\psi_\theta(t) = \begin{bmatrix} \eta_0^T(t) & 0 & 0 & 0 \\ 0 & \eta_0^T(t) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \eta_0^T(t) \end{bmatrix} \quad (12)$$

an  $n_y \times n_\theta$ , matrix

$$\theta_0 = [\sigma_1^T \quad \sigma_2^T \quad \dots \quad \sigma_{n_\theta}^T]^T \quad (13)$$

where

$$\begin{aligned} \eta_0^T(t) &= (-y(t-1)^T, \dots, -y(t-n_a)^T, u(t-1)^T, \dots, u(t-n_b)^T, e(t-1)^T, \dots, e(t-n_c)^T) \\ \sigma_i^T &= i\text{th row of } [A_1 \dots A_{n_a} \quad B_1 \dots B_{n_b} \quad C_1 \dots C_{n_c}] \end{aligned} \quad (14)$$

If (11) is now formally considered to be the observation equation of a state representation of system (8) with the state defined as the (constant) parameter vector  $\theta_0(t) = \theta_0$ , we obtain a particular form of state equations for (8):

$$\begin{aligned} \theta_0(t+1) &= \theta_0(t) \\ y^\theta(t) &= \psi_\theta(t) \theta_0(t) + e(t) \end{aligned} \quad (15)$$

For parameter estimation purposes, we also need a model of (15) based on parameter value  $\theta(t)$  :

$$\begin{aligned} \theta(t+1) &= \theta(t) \\ y^o(t) &= \psi(t)\theta(t) + e(t) \end{aligned} \quad (16)$$

where, in analogy with (11)-(14),

$$\psi(t) = \begin{bmatrix} \eta^T(t) & 0 & 0 & 0 \\ 0 & \eta^T(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \eta^T(t) \end{bmatrix} \quad (17)$$

is a matrix of dimension  $n_y \times n_\theta$ .

$$\theta(t) = [\sigma_1^T(t) \ \sigma_2^T(t) \ \cdots \ \sigma_n^T(t)]^T \quad (18)$$

where

$$\begin{aligned} \eta^T(t) &= (-y(t-1)^T, \dots, -y(t-n_a)^T, u(t-1)^T, \dots, u(t-n_b)^T, e(t-1)^T, \dots, e(t-n_c)^T) \\ \sigma_i^T &= \text{ith row of } [A_1(t) \cdots A_{n_a}(t) \ B_1(t) \cdots B_{n_b}(t) \ C_1(t) \cdots C_{n_c}(t)] \\ \varepsilon(t) &= y^o(t) - \psi(t)\theta(t), \ E\varepsilon(t)\varepsilon^T(t) = \Lambda \end{aligned} \quad (19)$$

By deriving the representation of the form (16), the parameter estimation problem is now transformed into a state estimation problem for (16). Let us now briefly comment on the structure of (16). While (15) is only of symbolic value since the random vectors  $e(t-1), \dots, e(t-n_c)$  in  $\psi_0(t)$  are unknown, these quantities are replaced in the observation matrix  $\psi(t)$  in (16) by the computed residuals  $\varepsilon(t-1), \dots, \varepsilon(t-n_c)$  as shown in (19), which makes (16) a usable model. The matrix  $\psi(t)$  now depends on  $\theta(\cdot)$ , and the representation (16) is nonlinear so that the UKF must be used. Note that the sequence  $\{e(t)\}$  is not white, but becomes white as  $\theta(t) \rightarrow \theta_0$ .

From the conversion of a parameter estimation problem into an UKF problem, we shall also derive some additional benefits. The available information about the statistical properties of the parameters to be estimated can be readily incorporated into the algorithm as has already been discussed, and the minimum variance Kalman filtering approach leads us to expect that lower variance of the parameter estimates can be obtained then from a number of other approaches available. Also, by a natural extension of the algorithm, it will be possible to estimate parameters which are Gauss-Markov random processes. What is left now is to apply the UKF algorithm to the state equation model (16). To make use of the minimum variance formulation of the Kalman filter, an estimator for the covariance  $\Lambda_0$  will be coupled with the filter equations. The resulting algorithm is particularly simple:

The n-dimensional random variable  $\theta(t)$  with mean  $\hat{\theta}(t)$  and covariance  $P^\theta(t)$  can be approximated by sigma points  $\chi_i(t)$  selected from the columns of  $\hat{\theta}(t) \pm (a\sqrt{LP^\theta(t)})_i$ ,  $i=0, \dots, 2L$ . The opposite weight  $\omega_i$  is  $\omega_0=1-(1/a^2)$ ,  $\omega_b(i)=1/2La^2$  ( $i=1, 2, \dots, 2L$ ).

The predicted mean and covariance are computed as

$$\hat{\theta}(t+1|t) = \sum_{i=0}^{2n} \omega_b(i) \chi_i(t) \quad (20)$$

$$P^\theta(t+1|t) = \sum_{i=0}^{2n} \omega_b(i) (\chi_i(t) - \hat{\theta}(t+1|t)) (\chi_i(t) - \hat{\theta}(t+1|t))^T \quad (21)$$

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

$$y_i^o(t+1|t) = \psi(t)\chi_i(t), \ \hat{y}^o(t+1|t) = \sum_{i=0}^{2n} \omega_b(i) y_i^o(t+1|t) \quad (22)$$

$$P_e^\theta(t+1|t) = \sum_{i=0}^{2n} \omega_b(i) (y_i^o(t+1|t) - \hat{y}^o(t+1|t)) (y_i^o(t+1|t) - \hat{y}^o(t+1|t))^T + \hat{\Lambda}(t+1) \quad (23)$$

$$P_{\theta y}^\theta(t+1|t) = \sum_{i=0}^{2n} \omega_b(i) (\chi_i(t+1) - \hat{\theta}(t+1|t)) (y_i^o(t+1|t) - \hat{y}^o(t+1|t))^T \quad (24)$$

$$K^\theta(t+1) = P_{\theta y}^\theta(t+1|t) [P_e^\theta(t+1|t)]^{-1} \quad (25)$$

$$\hat{\theta}(t+1) = \hat{\theta}(t+1|t) + K^\theta(t+1)(y^o(t+1) - \psi(t)\chi_i(t)), \ \hat{\theta}(0) = \theta^0 \quad (26)$$

$$P^\theta(t+1) = P^\theta(t+1|t) - K^\theta(t+1)P_e^\theta(t+1)[K^\theta(t+1)]^T, \ P(0) = \begin{bmatrix} \Pi(\hat{\theta}^0) & 0 \\ 0 & \Sigma^0 \end{bmatrix} \quad (27)$$

$$\hat{\Lambda}(t+1) = \hat{\Lambda}(t) + \frac{1}{t+1} [\varepsilon(t)\varepsilon^T(t) - \hat{\Lambda}(t)], \ \hat{\Lambda}(0) = \Lambda^0 \quad (28)$$

$$\varepsilon(t) = y^o(t) - \psi(t)\theta(t) \quad (29)$$

In (28), in analogy with (27),  $\hat{\Lambda}(t+1)$  denotes an estimate of  $\Lambda$  based on  $t+1$  data pairs  $\{u(j), y(j)\}$ ,  $j=0, 1, \dots, t$ , and the resulting formula (28) is obtained by rewriting its off-line version.

To avoid some numerical problems which may arise if the matrix  $P^\theta(t+1)$  becomes singular, (21) is sometimes replaced by

$$\hat{P}^\theta(t+1|t) = \sum_{i=0}^{2n} \omega_b(i) (\chi_i(t) - \hat{\theta}(t+1|t)) (\chi_i(t) - \hat{\theta}(t+1|t))^T + \Delta Q \quad (30)$$

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.

**Remark 1.** To ensure the convergence of UKF, the matrices  $P^\theta(t+1)$  need to be positive definite. From (27), as  $P^\theta(t+1|t)$  and  $-K^\theta(t+1)P_e^\theta(t+1)[K^\theta(t+1)]^T$  may be not positive definite matrices, extra additive matrix  $\Delta Q$  in (30) should be introduced as a modification to the UKF so that  $P^\theta(t+1) \geq 0$  will be satisfied. Obviously, if  $\Delta Q$  is sufficiently large,  $P^\theta(t+1)$  can always be positive definite. This means that the UKF can tolerate high order error introduced during the UT by enlarging the noise covariance matrix.

#### 4. CONVERGENCE ANALYSIS

In this section, a simple approach to convergence of the algorithm (20)-(30) is given.

Expanding  $\psi(t)\theta(t)$  in (16) by a Taylor series about  $\theta = \hat{\theta}(t)$  gives,

$$G(t) = \frac{\partial}{\partial \theta} \psi(t)\theta \Big|_{\theta=\hat{\theta}(t)} = - \frac{\partial}{\partial \theta} \varepsilon(t) \Big|_{\theta=\hat{\theta}(t)} \quad (31)$$

Substituting (31) into (16) yields an approximate equality.

$$y^o(t) \approx G(t)\theta(t) + e(t) \quad (32)$$

In (32), it is evident that there always exist residuals of error prediction  $\hat{\theta}(t+1|t)$ . In order to take these residuals into account and obtain a more exact equality, an unknown instrumental diagonal matrix  $\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$  is introduced, so that

$$y^o(t) = \lambda G(t)\theta(t) + e(t) \quad (33)$$

To prove the convergence of the algorithm (20)-(30), a method of analysis based on an associated deterministic differential equation (Ljung, 1977) will be used. To be able to apply the analysis, we shall assume in the following that the input  $u(t)$  is a weakly stationary process with rational spectral density and that all absolute moments of the sequences  $\{u(t)\}$ ,  $\{e(t)\}$  exist and are bounded, we shall also require that the system (8) and the equation generating  $\varepsilon(t)$  in (29) be stable.

For analysis of the algorithm convergence, some regularity conditions similarly as in (Ljung, 1977; Ljung, 1979; Xiong et al., 2006) are recalled.

**Lemma.** Consider the differential equation given by

$$\frac{d}{d\tau} \theta^D(\tau) = f(\theta^D(\tau)) \quad (34)$$

Let  $D_s = \{\theta | (A(\theta), Q) \text{ stabilizable and } (A(\theta), C(\theta)) \text{ detectable}\}$ .  
 Let  $\{\hat{\theta}(t), \hat{x}(t)\}$  be given by algorithm (20)-(30).

1) Suppose that the associated deterministic differential equation (34) has an invariant set  $D_c$ , with domain of attraction  $\theta \in D_A$  (which will be a subset of  $D_s$ ). Suppose further that  $\hat{\theta}(t)$  belongs to a compact subset of  $D_A$  and  $\hat{x}(t)$  is bounded infinitely often with probability one. Then

$$\hat{\theta}(t) \rightarrow D_c \text{ with probability 1 as } t \rightarrow \infty. \quad (35)$$

2) Suppose that  $\hat{\theta}(t) \rightarrow \theta^*$  with probability greater than zero. Then  $\theta^*$  must be a stable stationary point of the differential equation (34).

3) Let  $\bar{D}$  be a compact subset of  $D_s$  such that the trajectories of (34) that start in  $\bar{D}$  do not leave  $\bar{D}$ . Suppose that the estimates  $\hat{\theta}(t)$  are projected into  $\bar{D}$  and that (34) has an invariant set  $D_c$  with a domain of attraction  $D_A \supset \bar{D}$ . Then  $\hat{\theta}(t) \rightarrow D_c$  with probability 1 as  $t \rightarrow \infty$ .

To apply the Lemma and the theory in (Ljung, 1977), we shall start with a derivation of an alternative expression for the Kalman gain  $K(t)$  in (26). Writing (20) and (30) in terms of

$$\Theta(t) = [tP^\theta(t)]^{-1} \quad (36)$$

We obtain

$$\hat{\theta}(t+1) = \hat{\theta}(t) + (\lambda/t+1)\Theta^{-1}(t+1)G^T(t)\hat{\Lambda}^{-1}(t+1)[y^\theta(t) - \psi(t)\hat{\theta}(t)] \quad (37)$$

$$\Theta(t+1) = \Theta(t) + (1/t+1)[K^\theta(t+1)P_e^\theta(t+1)[K^\theta(t+1)]^T - \Theta(t)] \quad (38)$$

The associated differential equation for the algorithm (20)-(30) will now be defined in terms of processes resulting from the algorithm when the parameter estimates are kept at some constant value  $\hat{\theta}(t) = \theta$ ,  $\hat{\Lambda}(t) = \Lambda$ . Then  $G(t)$ ,  $\psi(t)$ , and  $\varepsilon(t)$  would give some  $\bar{G}(t; \theta)$ ,  $\bar{\psi}(t; \theta)$  and  $\bar{\varepsilon}(t; \theta)$  for large  $t$ . Define now the expectations

$$\begin{aligned} f(\theta, \Lambda) &= \lambda E \bar{G}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta) \\ W(\theta) &= E \bar{\varepsilon}(t; \theta) \bar{\varepsilon}^T(t; \theta) \end{aligned} \quad (39)$$

The behavior of the estimates obtained from (20)-(30) can be described by the coupled differential equations

$$\begin{aligned} d\theta(\tau) &= R^{-1}(\tau) f(\theta(\tau), \Lambda(\tau)) d\tau \\ d\Lambda(\tau) &= W(\theta(\tau)) - \Lambda(\tau) d\tau \end{aligned} \quad (40)$$

To comment on the significance of some of the quantities introduced in (39), the expression  $f(\theta, \Lambda)$  can be interpreted as the average correction to  $\hat{\theta}(t+1)$  in (37) with  $[1/t+1]\Theta^{-1}(t+1) = P(t+1)$  being the correction gain. Apparently,  $P(t+1) \rightarrow 0$  for large  $t$  if  $\Theta^{-1}(t+1)$  is bounded, but this is guaranteed by (38). The convergence properties of the algorithm (20)-(30) can now be summarized in the following result.

**Theorem.** Consider input-output data  $\{u(t), y(t) | t = 0, 1, \dots\}$  generated by a stable system (8), and its model (16) whose parameters are estimated by the algorithm (20)-(29) with (21) replaced by (30). Assume that the algorithm is equipped with a feature that guarantees stable generation of  $\varepsilon(t)$  in (29). (This is achieved by allowing only such parameter vector  $\hat{\theta}(t)$  for which  $\hat{C}^{-1}(q^{-1})$  is stable.) Then the estimate  $\hat{\theta}(t)$  converges with probability one to a stationary point of the function

$$V(\theta, \Lambda) = E \bar{\varepsilon}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta) + \ln |\Lambda| \quad (41)$$

where  $|\Lambda|$  denotes the determinant of  $\Lambda$  and

$$\bar{\varepsilon}(t, \theta) = C_\theta^{-1}(q^{-1}) A_\theta^{-1}(q^{-1}) y(t) - C_\theta^{-1}(q^{-1}) B_\theta^{-1}(q^{-1}) u(t)$$

where  $A_\theta$ ,  $B_\theta$  and  $C_\theta$  are model polynomials corresponding to  $\theta$ .

**Proof.** Under standing assumptions, the asymptotic properties of the algorithm (20)-(30) are described by the differential equation (40). From (29), (39), and (41) we get

$$f(\theta, \Lambda) = -\frac{1}{2} \frac{d}{d\theta} V(\theta, \Lambda)$$

Choosing now (41) as a Lyapunov function, this gives

$$\begin{aligned} \dot{V} &= E \frac{d}{d\theta} [\bar{\varepsilon}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta)] \Theta^{-1} f(\theta, \Lambda) \\ &\quad - E \bar{\varepsilon}^T(t; \theta) \Lambda^{-1} (W(\theta) - \Lambda) \bar{\varepsilon}(t; \theta) + \text{tr}[\Lambda^{-1} (W(\theta) - \Lambda)] \\ &= -2 f^T(\theta, \Lambda) \Theta^{-1} f(\theta, \Lambda) - \text{tr}[\Lambda^{-1} (W(\theta) - \Lambda)] \Lambda^{-1} (W(\theta) - \Lambda) \end{aligned} \quad (42)$$

The term  $-2 f^T(\theta, \Lambda) \Theta^{-1} f(\theta, \Lambda)$  on the right side of (42), with (39) it can be verified that

$$\begin{aligned} &-2 f^T(\theta, \Lambda) \Theta^{-1} f(\theta, \Lambda) \\ &\doteq -2 [\lambda E \bar{G}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta)]^T \Theta^{-1} [\lambda E \bar{G}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta)] \\ &= -2 \lambda^2 [E \bar{G}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta)]^T \Theta^{-1} [E \bar{G}^T(t; \theta) \Lambda^{-1} \bar{\varepsilon}(t; \theta)] \leq 0 \end{aligned}$$

The last term  $-\text{tr}[\Lambda^{-1} (W(\theta) - \Lambda)] \Lambda^{-1} (W(\theta) - \Lambda)$  on the right side of (42) was obtained by applying well-known trace identities, with (39) it is obviously that

$$-\text{tr}[\Lambda^{-1} (W(\theta) - \Lambda)] \Lambda^{-1} (W(\theta) - \Lambda) \leq 0$$

Therefore, we can obtain

$$\dot{V} = -2 f^T(\theta, \Lambda) \Theta^{-1} f(\theta, \Lambda) - \text{tr}[\Lambda^{-1} (W(\theta) - \Lambda)] \Lambda^{-1} (W(\theta) - \Lambda) \leq 0 \quad (43)$$

The equality in (43) holds only for such  $\theta = \theta^*$  and  $\Lambda = \Lambda^*$  which give  $f(\theta, \Lambda^*) = 0$  and  $\Lambda^* = W(\theta^*) = E \bar{\varepsilon}(t, \theta^*) \bar{\varepsilon}^T(t, \theta^*)$ . These are the stationary points of (41).  $\square$

**Remark 2.**  $\lambda$  is an unknown instrumental diagonal matrix introduced to evaluate the error introduced by linearization. And the convergence of the algorithm do not depend on the magnitude of  $\lambda$ . According to (41), although different  $\lambda$  may change the value of  $\dot{V}$  in (42), it will remain negative and the relationship shown in (43) will not be changed.

## 5. SIMULATION

The results in the preceding three sections clarify the UKF based nonlinear filtering for parameter estimation in linear systems with correlated noise and the convergence analysis of the algorithm, respectively.

In order to show the efficiency of the algorithm, it is applied to the GPS error model of the vehicle navigation systems with aided GPS (Nebot, 1998) in comparison with the UKF and the EKF.

In the GPS error model, the auto-correlation and corresponding power spectral density (PSD) of the error signal must be estimated. A transfer function can be fitted to the PSD estimates in the form

$$\Phi(s) = \left[ \frac{\gamma(s + \beta)}{s^2 + s2\alpha\beta + \alpha^2} \right]^2$$

with parameters unknown.

This model may be described by a shaping filter with process model

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} -2\zeta\alpha & 1 \\ -\alpha^2 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \gamma \\ \beta \end{bmatrix} w(t), \quad \zeta = 0.7$$

Then the constant velocity model and error model to estimate longitude and latitude is expressed as

$$\begin{aligned} x(t+1) &= Ax(t) + \rho(t) \\ z(t) &= Hx(t) + v(t) \end{aligned}$$

where

$$x(t) = [v_o(t) \quad v_a(t) \quad x_o(t) \quad x_a(t)]^T$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\zeta\alpha & 1 \\ 0 & 0 & -\alpha^2 & 0 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$Q = E[\rho(t)\rho^T(t)] = \begin{bmatrix} q_{vo}^2 & 0 & 0 & 0 \\ 0 & q_{va}^2 & 0 & 0 \\ 0 & 0 & \gamma^2 & \gamma^2\beta \\ 0 & 0 & \gamma^2\beta & (\gamma\beta)^2 \end{bmatrix}$$

$$R = E[v(t)v^T(t)] = \begin{bmatrix} \sigma_{gps}^2 & 0 \\ 0 & \sigma_{vel}^2 \end{bmatrix}$$

where,  $v_o(t)$  and  $v_a(t)$  are the velocity on the directions of longitude and latitude, respectively.  $x_o(t)$  and  $x_a(t)$  are the parameters of the variance of longitude and latitude, respectively.  $q_{vo}^2$ ,  $q_{va}^2$ ,  $\sigma_{gps}^2$  and  $\sigma_{vel}^2$  are white noises.

It is obviously that in this application the observation is composed of the GPS position measurement that is corrupted with correlated noise. Then the unknown parameters  $\alpha$ ,  $\beta$  and  $\gamma$  are represented as a parameter vector  $\theta$  which will be estimated by using the UKF approach and form an enlarged state vector by appending the parameter vector  $\theta = \theta(t)$  to the state. So we can obtain

$$\theta(t) = [\alpha \ \beta \ \gamma]^T, z(t) = [x(t) \ \theta(t)]^T$$

The required parameter estimates  $z(t)$  and  $\theta(t)$  are now obtained by applying the proposed UKF. The trajectories of the estimated parameters  $x_o(t)$  and  $x_a(t)$  with EKF and UKF are depicted in Fig. 1 and Fig. 2, respectively.

Fig. 1 and Fig. 2 show the filtering results of the system with respect to the GPS position estimate  $x_o(t)$  and  $x_a(t)$ . The GPS observation consists almost entirely of correlated noise because the vehicle is stationary. Compare with EKF, the position estimation with UKF is substantially obtained and the position error can remain in the 100 m range on the whole.

The trajectories of the estimated parameters  $\alpha$ ,  $\beta$  and  $\gamma$  with EKF and UKF are compared in Fig. 3 to Fig. 5 and with the new form of EKF and UKF are compared in Fig. 6 to Fig. 8, respectively.

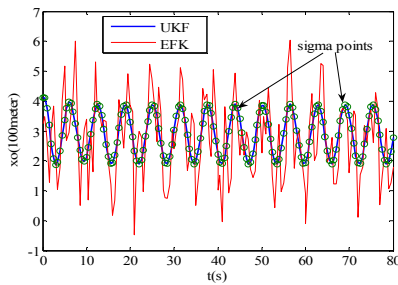


Fig. 1. The parameter estimation of the longitude  $x_o(t)$ .

Fig. 3 to Fig. 5 show the filtering results of the unknown parameters  $\alpha$ ,  $\beta$  and  $\gamma$  represented as a parameter vector  $\theta$  which are estimated by using the UKF approach. The filter is able to obtain the required parameter estimates  $\alpha$ ,  $\beta$  and  $\gamma$  for the GPS error model when the estimate parameters remain unknown. These results may be satisfied the anticipated values of GPS observation. From these figures, we can find UKF make sure the estimation achieve the anticipated values more quickly compare with EKF. But the computational burden is still substantial, though the divergence and bias problems don't appear.

Fig. 6 to Fig. 8 show the filtering results of the unknown parameters  $\alpha$ ,  $\beta$  and  $\gamma$  represented as a parameter vector  $\theta$

which are estimated by using the new proposed UKF approach. The filter also can obtain the required parameters and satisfy the anticipated values of GPS observation. And also, UKF make sure the estimation achieve the anticipated values more quickly compare with EKF. Furthermore, compare with Fig. 3 to Fig. 5, it is easy to see that the estimation achievement process to the anticipated values is expedited. Consequently, the computational burden is obviously reduced by using the new proposed UKF. From these figures, the convergence of the new proposed UKF is also verified.

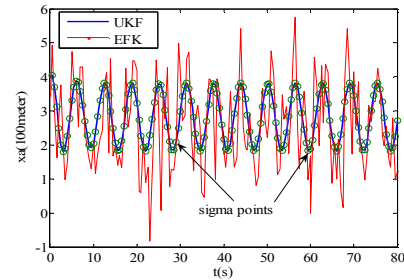


Fig. 2. The parameter estimation of latitude  $x_a(t)$ .

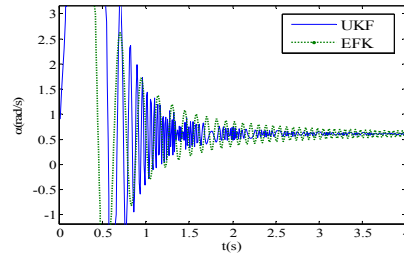


Fig. 3. The parameter  $\alpha$  estimation with UKF and EKF.

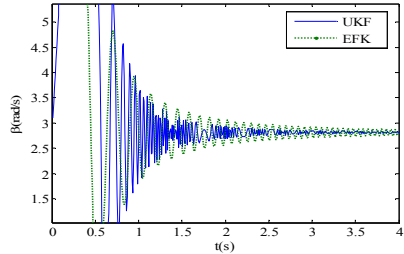


Fig. 4. The parameter  $\beta$  estimation with UKF and EKF.

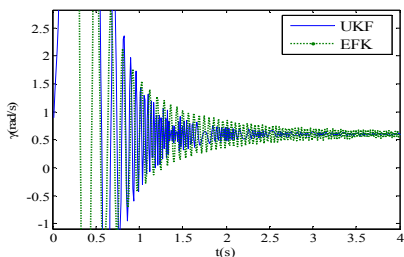


Fig. 5. The parameter  $\gamma$  estimation with UKF and EKF.

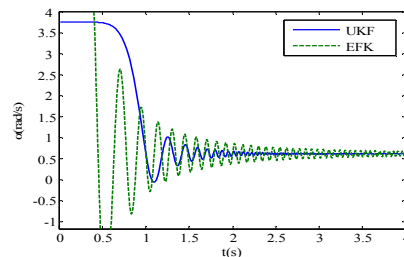


Fig. 6.  $\alpha$  estimation with the new form UKF and EKF.



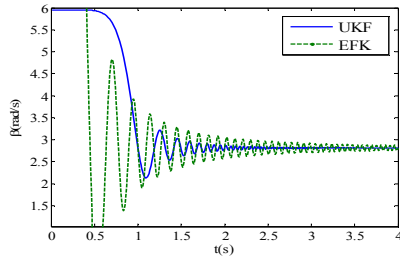


Fig. 7.  $\beta$  estimation with the new form UKF and EKF.

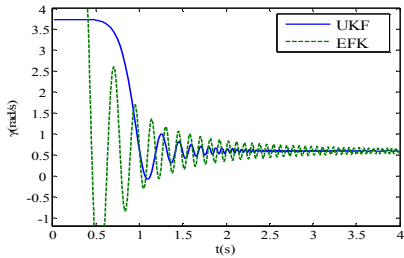


Fig. 8.  $\gamma$  estimation with the new form UKF and EKF.

The simulations on the vehicle navigation systems with aided GPS in this section verify the proposed algorithm and its performance from the view of experimentation. It is shown that the proposed algorithm has practicability to a certain extent.

## 6. CONCLUSIONS

Based on the nonlinear filter UKF, the algorithm is presented for linear systems with correlated noises. Firstly, the outline of the problem is presented and the UKF equations are proposed. Then the parameter estimation algorithm is developed and an extra additive matrix  $\Delta Q$  is introduced to make sure the matrices  $P(t+1)$  is positive definite. Furthermore, the analysis of the related convergence properties of the algorithm is given. According to some standard results about the convergence of stochastic processes, it is pointed out that, the convergence of the algorithm may be ensured and do not depend on the magnitude of  $\lambda$  which is an unknown instrumental diagonal matrix introduced to evaluate the error introduced by linearization. Moreover, the vehicle navigation systems with aided GPS are introduced to show the high performances of the proposed algorithm.

## ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China, under grant 60274009, and Specialized Research Fund for the Doctoral Program of Higher Education, under grant 20020145007.

## REFERENCES

Astrom K. J. and P. Eykhoff (1971). System identification - A survey. *Automatica*, **7**, 123-162.  
 Chin-Lang Tsai and Kao Wei-Wen (2006). Constrained total least-square solution for GPS compass attitude determination. *Applied Mathematics and Computation*, **183(1)**, 106-118.  
 Granados G. Seco and J. A. Ferncidez Rubio (1998). Maximum likelihood propagation-delay estimation in unknown correlated noise using antenna arrays:

application to global navigation satellite systems. *ICASSP '98*, **4**, 2065-2068.  
 James L. Garrison and Stephen J. Katzberg (2000). The Application of Reflected GPS Signals to Ocean Remote Sensing. *Remote Sensing of Environment*, **73(2)**, 175-187.  
 Jazwinski K. H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic.  
 Julier S. J., J. K. Uhlmann and H. F. Durrant-Whyte (1995). A new approach for filtering nonlinear systems. *Proc. of the 14<sup>th</sup> American Control Conference*, Seattle, Washington, 1628-1632. The AACC, Seattle, WA.  
 Julier S. J. and J. K. Uhlmann (1997a). A consistent, unbiased method for converting between polar and Cartesian coordinate systems. *The Proc of Aero Sense: 11th Int Symposium on Aerospace/ Defense Sensing, Simulation and Controls*, Orlando, 110-121.  
 Julier S. J. and J. K. Uhlmann (1997b). A new extension of the Kalman filter to nonlinear systems. *The Proc of Aero Sense: 11th Int Symposium Aerospace/ Defense Sensing, Simulation and Controls*, Orlando, 54-65.  
 Julier, S. J. and J. K. Uhlmann (1997c). A General Method for Approximating Nonlinear Transformations of Probability Distributions [EB/OL]. <http://www.Robots.ox.ac.uk/~siju/work/publications/Unscented.zip>, 1997-09-27.  
 Julier S. J. (2000). The scaled unscented transformation. *Proceedings of the American Control Conference*, Anchorage, AK, 4555-4559.  
 Julier S. J., J. K. Uhlmann and H. F. Durrant-Whyte (2000). A new approach for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. on Automatic Control*, **45(3)**, 477-482.  
 Julier S. J. and J. K. Uhlmann (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, **92(3)**, 401-422.  
 Lefebvre T., H. Bruyninckx and J. De Schutter (2002). Comment on 'a new method for the nonlinear transformation of means and covariances in filters and estimators'. *IEEE Trans on Automatic Control*, **47(8)**, 1406-1408.  
 Ljung L. (1979). Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. on Automatic Control*, **AC-24(2)**, 36-50.  
 Ljung L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. on Automatic Control*, **AC-22(8)**, 551-575.  
 Nebot E.M., H. Durrant Whyte and S. Scheding (1998). Frequency domain modeling of aided GPS for vehicle navigation systems. *Robotics and Autonomous Systems*, **25(1-2)**, 73-82.  
 Pan Quan, Yang Feng, Ye Liang, Lian Yan and Chen Yongme. (2005). Survey of a kind of nonlinear filters—UKF. *Control & Decision*, **20(5)**, 481-494 (in Chinese)  
 Wan E. A. and R. Van Der Merwe (2000). The unscented Kalman filter for nonlinear estimation. *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 153-158.  
 Xiong K., H. Y. Zhanga and C. W. Chan (2006). Performance evaluation of UKF-based nonlinear filtering. *Automatica*, **42(2)**, 261-270.