**Article**

**Design Method for a Higher Order Extended Kalman Filter Based on Maximum Correlation Entropy and a Taylor Network System**

Qiupeng Wang 1, Xiaohui Sun 2 and Chenglin Wen 3,*

1 School of HDU-ITMO, Joint Institute, Hangzhou Dianzi University, Hangzhou 310018, China; wangqp@hdu.edu.cn
2 School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China; sun_xh1993@hdu.edu.cn
3 School of Automation, Guangdong University of Petrochemical Technology, Maoming 525000, China
* Correspondence: wencl@hdu.edu.cn

**Abstract:** This paper proposes one new design method for a higher order extended Kalman filter based on combining maximum correlation entropy with a Taylor network system to create a nonlinear random dynamic system with modeling errors and unknown statistical properties. Firstly, the transfer function and measurement function are transformed into a nonlinear random dynamic model with a polynomial form via system identification through the multidimensional Taylor network. Secondly, the higher order polynomials in the transformed state model and measurement model are defined as implicit variables of the system. At the same time, the state model and the measurement model are equivalent to the pseudolinear model based on the combination of the original variable and the hidden variable. Thirdly, higher order hidden variables are treated as additive parameters of the system; then, we establish an extended dimensional linear state model and a measurement model combining state and parameters via the previously used random dynamic model. Finally, as we only know the results of the limited sampling of the random modeling error, we use the combination of the maximum correlation estimator and the Kalman filter to establish a new higher order extended Kalman filter. The effectiveness of the new filter is verified by digital simulation.

**Keywords:** pseudolinearization; maximum entropy Kalman filter; multidimensional Taylor network

---

**1. Introduction**

The application of filters occupies an important position in various fields at the national and international levels. The progress and development of filters play important roles in national economic construction—especially national defense construction—such as in real-time estimation and target tracking. In 1960, Kalman proposed a method of filtering under the minimum mean squared error criterion for linear systems, and it soon began to be widely used [1]. In order to solve nonlinear problems, extended Kalman filters (EKFs) [2], unscented Kalman filters (UKFs) [3], and cubature Kalman filters (CKFs) have since emerged. However, the above-mentioned filtering methods require the modeling error to be Gaussian white noise. As such, their performances are likely to worsen when applied to non-Gaussian situations, especially when the systems are disturbed by impulsive noise. Impulsive noise arises from heavy-tailed distributions [4] (such as some mixed Gaussian distributions), and is common in many real scenarios of automatic control and target tracking (for instance, the measurement noise in the radar system is often not Gaussian noise, but heavy-tailed non-Gaussian noise [5]). In 1993, Gordon and Salmond proposed particle filtering when the density function is known [6]; this achieves an approximation of the distribution function by sampling a large number of particles therein. However, this method is very complicated; it requires a large number of particles, and it will cause particle degradation after re-sampling. In general, the density function is difficult
to obtain. For this reason, for the linear system, Chen designed the corresponding Kalman filter under the maximum correlation entropy criterion based on the limited realization of random variables [7]; this is called the maximum correlation entropy Kalman filter (MCKF) [8]. On this basis, the maximum correntropy extended Kalman filter (MCEKF) and the maximum correntropy unscented Kalman filter (MCUKF), which can solve nonlinear non-Gaussian systems, have since emerged [9]. However, in MCEKFs, all higher order terms in the Taylor expansion are discarded. Therefore, a large truncation error will be generated, and the filtering performance will decrease or even diverge as the nonlinearity of the system increases. In addition, each step of the state estimation needs to recalculate the Taylor expansion coefficient, which will undoubtedly increase the complexity of the calculation. MCUKFs use UT transformation and sigma point sampling [10]; this is called deterministic sampling. There is only one sampling point for a dimensional system. Neither low-dimensional nor high-dimensional systems have a strong claim to superiority. A large number of experiments have shown that both EKFs and UKFs can be approximated by a second-order polynomial at most [11], which will produce a large rounding error. Hence, both will eventually face the problems of degraded filtering performance and divergence as their nonlinearity increases [12].

This project proposes a higher order extended Kalman filter method based on maximum correlation entropy, under the assumption that both state and measurement equations can be modeled and based on a strong nonlinear function. The main contributions of this paper are as follows: (1) using multidimensional Taylor nets to convert the general expression of nonlinear functions into higher order polynomials; (2) defining each order of polynomial in the system as hidden variables of the corresponding order, and treating them as time-variable parameters; (3) establishing the dynamic relationship between the time-variable parameters and combining them with the original variables to further establish the expanded dimension state model; (4) based on the expanded linear state variables, equivalently rewriting the measurement model into the corresponding linear form; and (5) according to the established linear state and measurement model of the new extended dimension system, establishing a higher order extended Kalman filter method based on maximum correlation entropy.

The remaining parts of this paper are organized as follows: the first chapter is the preface of our knowledge, which introduces the definition of “entropy”; the Section 2 presents a method for identifying nonlinear functions based on multidimensional Taylor networks; the Section 3 presents a higher order extended Kalman filter method; the Section 4 presents the detailed design process of the maximum correlation entropy higher order extended Kalman filter; the Section 5 concerns simulation verification; and the Sections 6 and 7 presents a summary and outlook.

2. Description of Correntropy

Correntropy is a generalized similarity measure between two random variables [13]. Given two one-dimensional random variables \( \varphi, \zeta \in \mathbb{R}^1 \), their joint distribution function is \( F_{\varphi, \zeta}(\varphi, \zeta) \); then, the correlation entropy is defined as follows:

\[
V(\varphi, \zeta) = \varepsilon[a(\varphi, \zeta)] = \int a(\varphi, \zeta)dF_{\varphi, \zeta}(\varphi, \zeta)
\]  

(1)

where \( \varepsilon \) is the expectation operator and \( a(\cdot, \cdot) \) is the translation-invariant Mercer kernel. In this article, it is not particularly emphasized that this kernel function is a Gaussian kernel, which is defined as follows:

\[
a(\varphi, \zeta) = G_T(\varepsilon) = \exp\left(-\frac{\varepsilon^2}{2\tau^2}\right)
\]  

(2)

where \( \varepsilon = \varphi - \zeta, \tau > 0 \) represents the kernel’s bandwidth.
By expanding Equation (2) with a Taylor series, we can obtain the following:

\[ a(\varphi, \zeta) = G_\tau(e) = \exp\left(-\frac{e^2}{2\tau^2}\right) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} \sum_{\tau=0}^{2k} \epsilon \{(\varphi - \zeta)^{2k}\} \]  

and then the correlation entropy of Equation (1) has the following expression:

\[ V(\varphi, \zeta) = \epsilon[a(\varphi, \zeta)] = \int \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} (\varphi - \zeta)^{2k} dF_{\varphi\zeta}(\varphi, \zeta) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} \sum_{\tau=0}^{2k} \epsilon \{(\varphi - \zeta)^{2k}\} \]  

where \( \epsilon \{(\varphi - \zeta)^{2k}\} = \int (\varphi - \zeta)^{2k} dF_{\varphi\zeta}(\varphi, \zeta) \) is the 2k truncation statistic of the random variable \( \varphi, \zeta \in R \).

However, in most practical cases, joint distribution \( F_{\varphi\zeta} \) is usually unknown, and there are often finite implementations \( (\varphi^{(j)}, \zeta^{(j)}), j = 1, 2, \cdots, N \) of \( (\varphi, \zeta) \) for random variables. In these cases, the sample mean estimator can be used to estimate the heterogeneity:

\[ \hat{\epsilon} \{(\varphi - \zeta)^{2k}\} = \frac{1}{N} \left( \sum_{j=0}^{N} (\varphi^{(j)} - \zeta^{(j)})^{2k} \right) \]  

Then, the entropy expression of the random variable pair \( (\varphi, \zeta) \) is driven by finite data:

\[ \hat{V}(\varphi, \zeta) = \hat{\epsilon}[a(\varphi, \zeta)] = \int \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} (\varphi - \zeta)^{2k} dF_{\varphi\zeta}(\varphi, \zeta) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} \sum_{\tau=0}^{2k} \hat{\epsilon} \{(\varphi - \zeta)^{2k}\} \]  

(6)

When \( \varphi, \zeta \in R^n \), and the components of vector \( e = \varphi - \zeta \) are independent of one another, multidimensional correlation entropy is based on N sampling.

3. Non-Linear Model Identification Based on Multidimensional Taylor Networks

Given that the state model and observation model are complex dynamic systems with nonlinear characteristics [14]:

\[ a(\tau + 1) = \sigma(a(\tau)) + \gamma(\tau) \]  

(7)

\[ \Gamma(\tau + 1) = \delta(a(\tau + 1)) + \theta(\tau + 1) \]  

(8)

where \( a(\tau) \in R^h \) is an h-dimensional state vector; \( T(\tau + 1) \in R^d \) represents the d-dimensional measure vector; and \( \sigma_i(a(\tau)), i = 1, 2, \cdots, h \) and \( \delta_j(a(\tau + 1)), j = 1, 2, \cdots, d \) represent the state function and the measurement function, respectively. The modeling errors for non-Gaussian systems are \( \gamma(k) \) and \( \theta(k) \), while \( \theta = \text{diag} \{ \theta_1, \theta_2, \cdots, \theta_b \} \) and \( \eta = \text{diag} \{ \eta_1, \eta_2, \cdots, \eta_d \} \) are the process noise variance and the measurement noise variance, respectively.

**Lemma 1.** Any continuous function defined in a closed interval can be approximated accurately with a polynomial function [15].
Lemma 2. For continuous functions, \( \sigma(\alpha(k)) \), defined in a closed interval, can be approximated by the following [16]:

\[
\sum_{i=1}^{N(h,l)} \psi_i(k) \prod_{t=1}^{l} \alpha_{i,t}(k)
\]

where \( N(h,l) \) denotes the total number of terms in the expansion and \( \lambda_{i,t} \) denotes the power of the variable \( \alpha_t \) in the product of the \( i \)th variable.

3.1. Multidimensional Taylor Network Structure

The multidimensional Taylor network model can replace the traditional neural network with the dynamic model and control the system under certain conditions; it is characterized by a nonlinear autoregressive moving-average model composed of polynomials. The multidimensional Taylor network (MTN) uses a forward single intermediate layer structure, including an input layer, an intermediate layer, and an output layer. Supposing that the input layer comprises \( n \) nodes—

\[
\alpha(\tau) = \begin{bmatrix} \alpha_1(\tau) & \alpha_2(\tau) & \cdots & \alpha_h(\tau) \end{bmatrix}^T \in \mathbb{R}^h
\]

the output layer is \( \alpha(\tau + 1) \), the middle layer is the network processing layer, and each input variable realizes the weighted summation of each power product term in this layer. The middle layer is composed of various power product terms and the corresponding connection weight vector \( \psi_j(\tau) \):

\[
\psi_j(\tau) = \begin{bmatrix} \psi_{j,1}(\tau), \psi_{j,2}(\tau), \cdots, \psi_{j,N(h,l)}(\tau) \end{bmatrix}^T
\]

which represents the output weight vector connecting the intermediate layer and the output node of the network.

According to the multivariate Taylor equation, if a function is differentiable to the \( h+1 \)th order at a certain point, then the function expands to a form where the power series of the variable is not greater than \( m \) times. The model can be expressed as a dynamic equation, as follows:

\[
\alpha_{j}(\tau + 1) = \sigma(\alpha(\tau)) = \sum_{i=1}^{N(h,l)} \psi_{j,i}(\tau) \prod_{t=1}^{l} \alpha_{i,t}^{\lambda_{i,t}}(\tau) + \Delta \sigma(\tau)
\]

where \( \sigma(\cdot) \) is a function of nonlinearity described by a multidimensional Taylor network model, \( \psi_j \) represents the weight before the product item of the \( i \)th variable, \( N(h,l) \) denotes the total number of terms in the expansion, \( \lambda_{i,t} \) denotes the power of the variable \( \alpha_t \) in the product of the \( i \)th variable, and \( \Delta \sigma(\tau) \) is the error—also known as the remainder—produced by the identification of a function by a multidimensional Taylor network.

3.2. Parameter Identification Method Based on Kalman Filtering

Model Establishment of a Kalman Filter

A Kalman filter can be regarded as an optimized autoregressive data processing method that describes the entire system through a state equation and an observation equation.

State equation:

\[
\psi_{j,i}(\tau + 1) = \psi_{j,i}(\tau) + w_{j,i}(\tau)
\]

where \( i = 1, 2, \cdots, N(h,l) \), \( j = 1, 2, \cdots, h \).

Observation equation:
It is not difficult to draw from Figure 1:

\[
\alpha_j(\tau + 1) = \sum_{i=1}^{N(h,l)} \psi_{ji,1}(\tau + 1) \prod_{l=1}^{j} \alpha_i(\tau + 1)
\]

\[
= H_j(\tau + 1) \cdot [\psi_{ji,1}(\tau + 1), \psi_{ji,2}(\tau + 1), \ldots, \psi_{ji,N(h,l)}(\tau + 1)]^T + \nu_j(\tau + 1)
\]

Thus,

\[
\alpha(\tau + 1) = [\alpha_1(\tau + 1), \alpha_2(\tau + 1), \ldots, \alpha_i(\tau + 1), \ldots, \alpha_h(\tau + 1)]^T
\]

(12)

where \( H(\tau) = [H_1(\tau + 1), H_2(\tau + 1), \ldots, H_h(\tau + 1)]^T; \psi(\tau + 1) = [\psi_1(\tau + 1), \psi_2(\tau + 1), \ldots, \psi_h(\tau + 1)]; \)

\( \psi_j(\tau + 1) \) represents the system state at \( \tau \)-time, that is, the parameter status value of the kth moment; and \( \beta(\tau + 1) \) represents the output value of the network. It is assumed that both process noise \( w(\tau) \) and \( v(\tau + 1) \) are Gaussian white noise during the analysis, and \( Q_j = \text{diag} \left( Q_{j,1}, Q_{j,2}, \ldots, Q_{j,N(h,l)} \right) \) and \( R_j = \text{diag} \left( R_{j,1}, R_{j,2}, \ldots, R_{j,N(h,l)} \right) \), which are the process noise variance and measurement noise variance, respectively. Here, we use a Kalman filter to approximate the dynamic model. As the filtering principle of Kalman fillters is mentioned later in this article, please refer to Equations (20)–(24) for the detailed process.

![Diagram of a multidimensional Taylor network](image)

**Figure 1.** Model of a multidimensional Taylor network.

### 3.3. Approximation Analysis

Given a class of nonlinear functions \( \sigma(\alpha(k)) \), it can be assumed that it is derivative of the rth order, but \( r \) is a relatively large number, making it difficult for us to use Taylor nets to approximate its function. The optimal approach would be to set \( m, 1 \leq m \leq r \) and use the Taylor network to expand the nonlinear function to the mth order, obtain the result of Equation (16), and simultaneously ensure the higher order error term \( \Delta \delta \leq \theta \), where \( \theta \) is the acceptable error threshold. This not only makes the Taylor network fitting function process easier, but also ensures the accuracy of the fit.
Sensors 2021, 21, 5864

4. Higher Order Extended Kalman Filter
4.1. Pseudolinearized Representation of Nonlinear Functions

For ease of description and understanding, if \( l = d = 2 \), we can expand Equation (7) through a multidimensional Taylor network to the \( m \)th order, as follows:

\[
\sigma_i(\alpha(k)) = (\alpha_{1,1,0}^1(\tau) + \alpha_{0,1,0}^2(\tau))
+ (\omega_{1,1,0}^1(\tau)^2 + \alpha_{1,1,1}^1(\tau)\alpha_2^2(\tau) + \alpha_{0,0,2}^2(\tau)^2)
+ (\alpha_{1,0,2}^1(\tau)^3 + \alpha_{1,1,2}^1(\tau)\alpha_2^2(\tau) + \alpha_{0,2,0}^2(\tau)^3)
+ \sum_{l_1 + l_2 + l_3 \leq l} \omega_{1,1,2}^1(\tau)\omega_{2}^2(l_1) + \cdots + \sum_{m_1 + m_2 \leq r} \omega_{l_1,l_2}^1(\tau)\omega_{m_2}^2(l_1) + \Delta \sigma_i(\tau)
\]

where \( \sum \omega_{l_1}^1(\tau)\omega_{l_2}^2(l_1) \) is the sum of all tensors of the \( l \)th order and \( \omega_{l_1,l_2}^1 \) represents the weight corresponding to each order of the tensor.

Definition 1. \( \alpha^{(l)}(\tau) = \left\{ \alpha_{1}^{(l)}(\tau), \alpha_{2}^{(l)}(\tau), \ldots, \alpha_{h}^{(l)}(\tau), \quad l_1 + l_2 + l_3 = l; \quad 0 \leq l \leq 6; \quad l = 0, 1, \ldots, h \right\} \) is a set of implicit variables of the \( l \)th order.

Definition 2. \( \omega_{l_1}^{(i)} = \left[ \omega_{l_1,1}^{(i)}, \omega_{l_1,2}^{(i)}, \ldots, \omega_{l_1,n_l}^{(i)} \right] = [\omega_{i,1,0}, \omega_{i,1,1}, \ldots, \omega_{i,n_l,0}] \), \( i = 1, 2, \ldots, l \) is the weight vector corresponding to the \( i \)th order implicit variable.

In [17], there is a detailed pseudolinearization process, so we will not repeat it in this article. In order to make the model more accurate, we treat the remainder \( \Delta \sigma(\tau) \) of the equation of state as latent variables. According to Definition 1 and Definition 2, the pseudolinear extended dimension form using the remainder as a hidden variable is as follows:

\[
\alpha^{(1)}(\tau + 1) = \mathcal{W}^{(1)}(\tau + 1, \tau)\alpha^{(1)}(\tau) + \sum_{i=2}^{m} \mathcal{W}^{(i)}(\tau + 1, \tau)\alpha^{(i)}(\tau) + C \cdot \Delta \sigma(\tau) + \gamma^{(1)}(\tau)
\]

where \( \alpha^{(i)}(\tau) = \left[ \alpha_{1}^{(i)}(\tau), \alpha_{2}^{(i)}(\tau) \right], \mathcal{W}^{(i)} = \left[ \begin{array}{c} \alpha_{1}^{(i)}(\tau) \\ \alpha_{2}^{(i)}(\tau) \end{array} \right], \gamma(\tau) = \gamma^{(1)}(\tau) = \left[ \begin{array}{c} \gamma_{1}^{(1)}(\tau) \\ \gamma_{2}^{(1)}(\tau) \end{array} \right], C = \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] \).

Similarly, Equation (8) can be rewritten as follows:

\[
\Gamma^{(1)}(\tau + 1) = \chi^{(1)}(\tau + 1)\alpha^{(1)}(\tau + 1) + \sum_{i=2}^{m} \chi^{(i)}(\tau + 1)\alpha^{(i)}(\tau + 1) + D \cdot \Delta \sigma(\tau + 1) + \theta^{(1)}(\tau + 1)
\]

where \( \Gamma^{(1)}(\tau + 1) = \left[ \Gamma_{1}^{(1)}(\tau), \Gamma_{2}^{(1)}(\tau) \right], \chi^{(i)} = \left[ \begin{array}{c} \chi_{1}^{(i)}(\tau) \\ \chi_{2}^{(i)}(\tau) \end{array} \right], \theta^{(1)}(\tau + 1) = \left[ \begin{array}{c} \theta_{1}^{(1)}(\tau + 1) \\ \theta_{2}^{(1)}(\tau + 1) \end{array} \right], D = \left[ \begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \right] \).

4.2. Linearized Representation of Nonlinear Functions

In order to transform the pseudolinear model established in Section 3.1 into a true linear form, it is necessary to establish a dynamic relationship between the \( l \)th order hidden variables and the \( u \)th order hidden variables [18]:

\[
\alpha^{(l)}(\tau + 1) = \mathcal{W}_{l}^{(u)}(\tau)\alpha^{(u)}(\tau) \quad l, u = 2, 3, \ldots, m
\]

where \( \mathcal{W} \) can be identified based on the multidimensional Taylor network in its original state; without any prior information, it can be set as follows:

\[
\mathcal{W}_{l}^{(u)}(\tau) = \left\{ \begin{array}{ll} 1, & l = u \\ 0, & l \neq u \end{array} \right.
\]
Combining Definition 1, Definition 2, and Equation (19), the state model Equation (7) has the following linear matrix form:

\[
W(\tau + 1, \tau) = \begin{bmatrix}
W_1^{(1)}(\tau) & W_1^{(2)}(\tau) & \cdots & W_1^{(m-1)}(\tau) & W_1^{(m)}(\tau) & C \\
W_2^{(1)}(\tau) & W_2^{(2)}(\tau) & \cdots & W_2^{(m-1)}(\tau) & W_2^{(m)}(\tau) & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
W_m^{(1)}(\tau) & W_m^{(2)}(\tau) & \cdots & W_m^{(m-1)}(\tau) & W_m^{(m)}(\tau) & 0 \\
0 & 0 & \cdots & 0 & 0 & C
\end{bmatrix}, \gamma(\tau) = \begin{bmatrix}
\gamma^{(1)}(\tau) \\
\gamma^{(2)}(\tau) \\
\vdots \\
\gamma^{(m-1)}(\tau) \\
\gamma^{(m)}(\tau)
\end{bmatrix},
\]

then, Equation (7) has the following linearized form:

\[
A(\tau + 1) = W(\tau + 1, \tau)A(\tau) + \gamma(\tau)
\]

where \(\gamma(k)\) is the modeling error.

In the same way, the linear matrix form of the measurement model can be obtained:

\[
\Gamma(\tau + 1) = \chi(\tau + 1, \tau)A(\tau + 1) + \theta(\tau + 1)
\]

where \(\chi(\tau + 1, \tau) = \begin{bmatrix}
\chi_1^{(1)}(\tau + 1) & \chi_2^{(1)}(\tau + 1) & \cdots & \chi_{m-1}^{(1)}(\tau + 1) & \chi_m^{(1)}(\tau + 1) & 0 & 0 \\
\chi_1^{(2)}(\tau + 1) & \chi_2^{(2)}(\tau + 1) & \cdots & \chi_{m-1}^{(2)}(\tau + 1) & \chi_m^{(2)}(\tau + 1) & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
\chi_1^{(m)}(\tau + 1) & \chi_2^{(m)}(\tau + 1) & \cdots & \chi_{m-1}^{(m)}(\tau + 1) & \chi_m^{(m)}(\tau + 1) & 0 & 0
\end{bmatrix},
\]

\[
\Gamma(\tau + 1) = \begin{bmatrix}
\Gamma_1(\tau) \\
\Gamma_2(\tau)
\end{bmatrix}, \text{ and } \theta(\tau + 1) = \begin{bmatrix}
\theta_1(\tau + 1) \\
\theta_2(\tau + 1)
\end{bmatrix}
\]
is the modeling error.

### 4.3. Design of Higher Order Extended Kalman Filter

For linear models, KF-based filters are given. Given the initial value \(A(0)\), when \(\gamma(\tau)\) and \(\theta(\tau + 1)\) are Gaussian white noise with zero mean, the variances are recorded as \(\theta\) and \(\eta\), respectively.

A recursive filter can be designed as follows:

\[
\hat{\lambda}(\tau + 1 | \tau) = W(\tau + 1, \tau)\lambda(\tau | \tau)
\]

\[
\lambda(\tau + 1 | \tau) = W(\tau + 1, \tau)\lambda(\tau | \tau)W^T(\tau + 1, \tau) + \theta(\tau)
\]

\[
K(\tau + 1) = (\lambda(\tau + 1 | \tau)\chi^T(\tau + 1)) (\chi(\tau + 1)\lambda(\tau + 1 | \tau)\chi^T(\tau + 1) + \eta(\tau + 1))^{-1}
\]

\[
\hat{\lambda}(\tau + 1 | \tau + 1) = \hat{\lambda}(\tau + 1 | \tau) + K(\tau + 1)((\Gamma(\tau + 1) - \chi(\tau + 1)\hat{\lambda}(\tau + 1 | \tau))
\]

\[
\lambda(\tau + 1 | \tau + 1) = (I - K(\tau + 1)\chi(\tau + 1))\lambda(\tau + 1 | \tau)
\]

### 5. Higher Order Extended Kalman Filter Design Based on Maximum Correlation Entropy

#### 5.1. Non-Gaussian Modeling of State Vector Based on Multivariate Information Observation

System status \(A(\tau)\) estimates \(\hat{\lambda}(\tau | \tau)\) and estimated error covariance \(\lambda(\tau | \tau)\) are obtained based on K [19]. The filtering equation predicts a step prediction estimation value \(\hat{\lambda}(\tau + 1 | \tau)\) and corresponding step prediction error covariance matrix \(\lambda(k + 1|k)\).
The step prediction estimate error of system status $A(\tau + 1)$ is as follows:

$$
\hat{A}(\tau + 1|\tau) = A(\tau + 1) - \hat{A}(\tau + 1|\tau)
$$

and it can be modified into a measurement model for system status $A(\tau + 1)$ as follows:

$$
\hat{A}(\tau + 1|\tau) = A(\tau + 1) - \hat{A}(\tau + 1|\tau)
$$

where $\hat{A}(\tau + 1|\tau)$ is a measurement of system status $A(\tau + 1)$, while $\hat{A}(\tau + 1|\tau)$ is the measurement error. Finally, the combined measurement model is as follows:

$$
\begin{bmatrix}
\hat{A}(\tau + 1|\tau) \\
\Gamma(\tau + 1)
\end{bmatrix} = 
\begin{bmatrix}
I \\
\chi(\tau + 1)
\end{bmatrix}A(\tau + 1) + \omega^{(j)}(\tau + 1)
$$

where $I$ is a unit array for the corresponding dimension, $\omega^{(j)}(\tau + 1)$ is received as follows:

$$
\begin{bmatrix}
\chi(\tau + 1)
\end{bmatrix}A(\tau + 1)
$$

where $\chi(\tau + 1)$ is the one-dimensional non-Gaussian vector and its components are not statistically independent. The non-Gaussian modeling error vector, and its components are not statistically independent. According to Equation (20), a step prediction error covariance of the system state $\lambda(\tau + 1)$ is received as follows:

$$
\tilde{\lambda}(\tau + 1|\tau) = W(\tau + 1, \tau)\lambda(\tau|\tau)W^T(\tau + 1, \tau) + \theta(\tau)
$$

where $\theta(\tau) = \text{diag}\{\theta^{(1)}(\tau), \theta^{(2)}(\tau), \cdots, \theta^{(m)}(\tau)\}$ and the $\theta^{(1)}(\tau), \theta^{(2)}(\tau), \cdots, \theta^{(m)}(\tau)$ is a covariance matrix of random error vectors $\gamma^{(1)}(\tau), \gamma^{(2)}(\tau), \cdots, \gamma^{(m)}(\tau)$ when the higher order hidden variable $\alpha^{(2)}(\tau), \cdots, \alpha^{(m)}(\tau)$ is dynamically modeled. $\theta^{(1)}(\tau)$ is the original system status model (Equation (16)) of the non-Gaussian model error $w^{(1)}(k)$, and calculates the second-order statistic after obtaining a limited number of samples:

$$
\tilde{\theta}^{(1)}(\tau) = \frac{1}{N}\sum_{j=1}^{N}\left\{[\gamma^{(1)}(\tau) - \overline{\gamma}(\tau)][\gamma^{(1)}(\tau) - \overline{\gamma}(\tau)]^T\right\}
$$

In Equation (23), $\tilde{\eta}(\tau + 1)$ is the calculated second-order statistic calculated after the non-Gaussian model error $\theta(\tau + 1)$ in the original system measurement state model (Equation (8)), obtaining a limited sample:

$$
\tilde{\eta}(\tau + 1) = \frac{1}{N}\sum_{j=1}^{N}\left\{[\theta^{(j)}(\tau + 1) - \overline{\theta}(\tau + 1)][\theta^{(j)}(\tau + 1) - \overline{\theta}(\tau + 1)]^T\right\}
$$

where $\theta^{(j)}(\tau + 1)$ is the $j$th realization vector of the non-Gaussian random noise vector $\theta(\tau + 1)$.

5.2. The Statistical Independence Process of Each Component in the Non-Gaussian Modeling Error Vector $\omega(\tau + 1)$ in the Comprehensive Measurement Model

The vector $\omega(\tau + 1)$ in the comprehensive measurement model Equation (22) is a non-Gaussian modeling error vector, and its components are not statistically independent. In order to use the correlation entropy form of the multidimensional independent vector shown in Equation (19), the one-dimensional non-Gaussian vector $\omega(\tau + 1)$ needs to be transformed into statistical independence.
From $\lambda(\tau + 1|\tau) = E\left\{\left[\Lambda(\tau + 1|\tau) - A(\tau|\tau)\right]\left[\Lambda(\tau + 1|\tau) - A(\tau|\tau)\right]^T\right\}$, $\lambda(\tau + 1|\tau)$ is a positive definite matrix. Similarly, in Equation (26), $\bar{\eta}(\tau + 1)$ is also a positive definite matrix. For this reason, Equation (23) is further expressed as follows:

$$
E\left\{\omega^{(i)}(\tau + 1)(\omega^{(i)})^T(\tau + 1)\right\} = \begin{bmatrix}
\Lambda_\alpha(\tau + 1|\tau)\Lambda_\alpha^T(\tau + 1|\tau) & 0 \\
0 & \Lambda_\Gamma(\tau + 1)\Lambda_\Gamma^T(\tau + 1)
\end{bmatrix}
$$

where $\Lambda_\alpha(\tau + 1)$ and $\Lambda_\Gamma(\tau + 1)$ are the Cholesky factor matrices of $\bar{\lambda}(\tau + 1|\tau)$ and $\bar{\eta}(\tau + 1)$, respectively.

Applying $\Lambda^{-1}(\tau + 1)$ to both sides of Equation (22), respectively, yields:

$$
\Lambda^{-1}(\tau + 1)\begin{bmatrix}
\hat{A}(\tau + 1|\tau) \\
\Gamma(\tau + 1)
\end{bmatrix} = \Lambda^{-1}(\tau + 1)\begin{bmatrix}
I \\
\chi(\tau + 1)
\end{bmatrix}A(\tau + 1) + \Lambda^{-1}(\tau + 1)\omega^{(i)}(\tau + 1)
$$

where

$$
D(\tau + 1) = \Lambda^{-1}(\tau + 1)\begin{bmatrix}
\hat{A}(\tau + 1|\tau) \\
\Gamma(\tau + 1)
\end{bmatrix},
S(\tau + 1) = \Lambda^{-1}(\tau + 1)\begin{bmatrix}
I \\
\chi(\tau + 1)
\end{bmatrix}, e(\tau + 1) = \Lambda^{-1}(\tau + 1)\omega(\tau + 1)
$$

The above equation can be further simplified as follows:

$$
D(\tau + 1) = S(\tau + 1)A(\tau + 1) + e(\tau + 1)
$$

because

$$
E\left\{e(\tau)e^T(\tau)\right\} = E\left\{\left[\Lambda^{-1}(\tau + 1)\omega(\tau + 1)\right][\Lambda^{-1}(\tau + 1)\omega(\tau + 1)]^T\right\}
$$

$$
= \Lambda^{-1}(\tau + 1)E\left\{\omega(\tau + 1)\omega^T(\tau + 1)\right\}(\Lambda^{-1}(\tau + 1))^T
$$

$$
= \Lambda^{-1}(\tau + 1)\Lambda(\tau + 1)\Lambda^T(\tau + 1)(\Lambda^{-1}(\tau + 1))^T
$$

Therefore, after the non-Gaussian modeling error random variable $\omega(\tau + 1)$ undergoes the equivalent transformation of the matrix $\Lambda^{-1}(\tau + 1)$, the components of the random $e(\tau + 1)$ are statistically independent.

5.3. Implementation Process of a Higher Order Extended Kalman Filter Based on Maximum Entropy

The filtering process of the extended Kalman filter (H-MCEKF) based on the maximum correlation entropy is as follows (see [20] for the specific derivation process):

1. The filter initialization obtains the initial filter value $\hat{A}(0)$ and the covariance $\lambda(0)$, choosing a suitable core bandwidth and a small positive number $\varepsilon$;
2. Taylor networks are used for system identification to obtain the parameters in the equations, using the expanded item and the remainder as the new hidden variables. A pseudolinearization process is performed to obtain the pseudolinear form of the system;
3. Equations (20) and (21) are used to obtain $\hat{X}(k + 1|k)$ and $P(k + 1|k)$, respectively, while Cholesky decomposition is used to obtain $P_\psi(k + 1|k)$;
4. $t = 1$ and $\hat{A}(\tau + 1|\tau + 1) = \hat{A}(\tau + 1|\tau)$ are taken, where $\hat{A}(\tau + 1|\tau + 1)$ represents the estimated state of the fixed-point iteration $t$;
5. The starting fixed-point iterative algorithm is as follows:

\[ \tilde{e}_i(\tau + 1) = d_i(\tau + 1) - s_i(\tau + 1)\hat{A}(\tau + 1 | \tau + 1)_{t-1} \]  

(36)

where \( e_i \) is the \( i \)th element of \( e \):

\[ \tilde{C}_a(\tau + 1) = diag(G(\tilde{e}_1(\tau + 1)), \ldots, G(\tilde{e}_m(\tau + 1))) \]  

(37)

\[ \tilde{C}_r(\tau + 1) = diag(G(\tilde{e}_{r+1}(\tau + 1)), \ldots, G(\tilde{e}_{m+m}(\tau + 1))) \]  

(38)

\[ \tilde{H}(\tau + 1) = \Lambda_r(\tau + 1)\tilde{C}_r^{-1}(\tau + 1)\Lambda_r^T(\tau + 1) \]  

(39)

\[ \tilde{\lambda}(\tau + 1 | \tau) = \Lambda_a(\tau + 1 | \tau)\tilde{C}_a^{-1}(\tau + 1)\Lambda_a^T(\tau + 1 | \tau) \]  

(40)

\[ \tilde{K}(\tau + 1) = \tilde{\lambda}(\tau + 1 | \tau)\chi^T(\chi\tilde{\lambda}(\tau + 1 | \tau)\chi^T + \tilde{H}(\tau + 1))^{-1} \]  

(41)

\[ \tilde{A}(\tau + 1 | \tau + 1)_t = \tilde{A}(\tau + 1 | \tau + 1) + \tilde{K}(\tau + 1)(\Gamma(\tau + 1) - \tilde{H}(\tau + 1)) \]  

(42)

The estimates of the current iteration step are compared with those of the previous iteration and, if satisfied,

\[ \| \tilde{A}(\tau + 1 | \tau + 1)_t - \tilde{A}(\tau + 1 | \tau + 1)_{t-1} \| \leq \varepsilon \]  

(43)

then \( \tilde{A}(\tau + 1 | \tau + 1) = \tilde{A}(\tau + 1 | \tau + 1)_t, \lambda(\tau + 1 | \tau + 1) = \tilde{\lambda}(\tau + 1 | \tau), \) and the value of the pseudovariable can be updated, or the iteration can be repeated;

6. \( \tau = \tau + 1, \) and steps (3–5) are repeated until the end of filtering.

6. Simulated Cases

This section verifies the validity of the proposed method by providing two cases: one in which the state equation is a nonlinear equation and the measurement equation is a linear equation, and one in which the state and measurement equations are both nonlinear.

6.1. Case 1

Consider a nonlinear system in which the state equation is a nonlinear model and the measurement equation is a linear model:

\[
\begin{align*}
    x_1(k + 1) &= 0.8 - 0.5e^{-x_1^2(k)/(1 + e^{-0.01k})} x_1(k) - (0.3 + 0.9e^{-x_1^2(k)/(1 + 0.5\sin(\pi k))}) x_2(k) + w_1(k) \\
    x_2(k + 1) &= 1.2(1 - e^{-0.8k}) x_2(k) + 0.11 x_1(k) + \cos(1 + x_2^2(k)) + e^{-0.8k} x_1^A(k) + w_2(k) \\
    y_1(k + 1) &= x_1(k + 1) + v_1(k + 1) \\
    y_2(k + 1) &= x_2(k + 1) + v_2(k + 1)
\end{align*}
\]

where the initial value \( x(0) \) is a random value of \([0, 1]\), the initial estimation error covariance \( P(0|0) = 0.1 \times diag(1, 1) \), and the process noise and measurement noise have the following characteristics:

\[
\begin{align*}
    w_1(k) &\sim 0.9N(0, 0.01) + 0.1N(0, 0.2), w_2(k) \sim 0.9N(0, 0.02) + 0.1N(0, 0.2) \\
    v_1(k) &\sim 0.9N(0, 0.01) + 0.1N(0, 2), v_2(k) \sim 0.9N(0, 0.02) + 0.1N(0, 2)
\end{align*}
\]

Figure 2 shows a diagram of the MTN identification system, while Figure 3 shows the estimated values of state variables \( x_1 \) and \( x_2 \) under the three filtering methods. From [21], we know the influence of \( \varepsilon \) is not significant compared with the kernel bandwidth \( \sigma \). The parameters are set at \( \varepsilon = 10^{-6} \). Tables 1 and 2 show the mean squared error and the mean relative error, respectively, of the estimated values under the three algorithms, which are computed as averages over 100 independent Monte Carlo runs, with each run containing 50 time steps. When \( \sigma = 5 \), the three algorithms all obtain better filtering results. Figures 4 and 5 show the probability densities of the estimation errors when estimating
the states $x_1$ and $x_2$, respectively, when the parameters are $\varepsilon = 10^{-6}$ and $\sigma = 5$. All of the results confirm that the proposed H-MCKF (design method for a higher order extended Kalman filter based on maximum correlation entropy and a Taylor network system) can outperform the MCEKF (maximum correntropy extended Kalman filter) significantly when the system is disturbed by non-Gaussian processes and measurement noise, and the H-MCKF_R (H-MCKF with the remainder of the state equation) further improves the filtering performance of the H-MCKF.

Table 1. The mean squared error using the three methods in case 1.

| $\sigma$ | $\varepsilon$ | MCEKF | H-MCKF | H-MCKF_R | MCEKF | H-MCKF | H-MCKF_R |
|---------|---------------|-------|--------|----------|-------|--------|----------|
| $2\sigma$ | $6 = 10^{-6}$ | 0.2073 | 0.1815 | 0.1694 | 0.1000 | 0.0774 | 0.0738 |
| $5\sigma$ | $6 = 10^{-6}$ | 0.1974 | 0.1244 | 0.1225 | 0.1100 | 0.0964 | 0.0921 |
| $10\sigma$ | $6 = 10^{-6}$ | 0.2282 | 0.1669 | 0.1636 | 0.1158 | 0.0925 | 0.0888 |
| $20\sigma$ | $6 = 10^{-6}$ | 0.2244 | 0.1602 | 0.1572 | 0.1160 | 0.0916 | 0.0880 |

Table 2. The mean relative error using the three methods in case 1.

| $\sigma$ | $\varepsilon$ | MCEKF | H-MCKF | H-MCKF_R | MCEKF | H-MCKF | H-MCKF_R |
|---------|---------------|-------|--------|----------|-------|--------|----------|
| $2\sigma$ | $6 = 10^{-6}$ | 0.3372 | 0.2405 | 0.2403 | 0.2354 | 0.2202 | 0.2084 |
| $5\sigma$ | $6 = 10^{-6}$ | 0.3462 | 0.2953 | 0.2906 | 0.2679 | 0.2485 | 0.2448 |
| $10\sigma$ | $6 = 10^{-6}$ | 0.3658 | 0.3052 | 0.2986 | 0.2745 | 0.2469 | 0.2426 |
| $20\sigma$ | $6 = 10^{-6}$ | 0.3634 | 0.3009 | 0.2945 | 0.2753 | 0.2466 | 0.2419 |

Figure 2. Graph of the MTN identification system in case 1.

Figure 3. (a,b) Set parameters: $\sigma = 5, \varepsilon = 10^{-6}$ in case 1; (c,d) set parameters: $\sigma = 10, \varepsilon = 10^{-6}$ in case 1.
Table 1. The mean squared error using the three methods in case 1.

| σ  | ε    | MSE of x₁ | MSE of x₂ |
|----|------|-----------|-----------|
| 2  | $10^{-6}$ | 0.2073    | 0.1815    |
| 2  | $10^{-6}$ | 0.1974    | 0.1244    |
| 2  | $10^{-6}$ | 0.2282    | 0.1669    |
| 2  | $10^{-6}$ | 0.2244    | 0.1602    |
| 5  | $10^{-6}$ | 0.1694    | 0.1225    |
| 5  | $10^{-6}$ | 0.1974    | 0.1244    |
| 5  | $10^{-6}$ | 0.2073    | 0.1815    |
| 5  | $10^{-6}$ | 0.1974    | 0.1244    |
| 5  | $10^{-6}$ | 0.2282    | 0.1669    |
| 5  | $10^{-6}$ | 0.2244    | 0.1602    |
| 10 | $10^{-6}$ | 0.1694    | 0.1158    |
| 10 | $10^{-6}$ | 0.1974    | 0.1244    |
| 10 | $10^{-6}$ | 0.2073    | 0.1815    |
| 10 | $10^{-6}$ | 0.1974    | 0.1244    |
| 10 | $10^{-6}$ | 0.2282    | 0.1669    |
| 10 | $10^{-6}$ | 0.2244    | 0.1602    |
| 20 | $10^{-6}$ | 0.1694    | 0.1158    |
| 20 | $10^{-6}$ | 0.1974    | 0.1244    |
| 20 | $10^{-6}$ | 0.2073    | 0.1815    |
| 20 | $10^{-6}$ | 0.1974    | 0.1244    |
| 20 | $10^{-6}$ | 0.2282    | 0.1669    |
| 20 | $10^{-6}$ | 0.2244    | 0.1602    |

Table 2. The mean relative error using the three methods in case 1.

| σ  | ε    | MRE of x₁ | MRE of x₂ |
|----|------|-----------|-----------|
| 2  | $10^{-6}$ | 0.3372    | 0.2405    |
| 2  | $10^{-6}$ | 0.3462    | 0.2953    |
| 2  | $10^{-6}$ | 0.3658    | 0.3052    |
| 2  | $10^{-6}$ | 0.3634    | 0.3009    |
| 5  | $10^{-6}$ | 0.2354    | 0.2202    |
| 5  | $10^{-6}$ | 0.2679    | 0.2485    |
| 5  | $10^{-6}$ | 0.2745    | 0.2469    |
| 5  | $10^{-6}$ | 0.2753    | 0.2466    |
| 10 | $10^{-6}$ | 0.2100    | 0.1964    |
| 10 | $10^{-6}$ | 0.2485    | 0.2202    |
| 10 | $10^{-6}$ | 0.2469    | 0.2485    |
| 10 | $10^{-6}$ | 0.2466    | 0.2485    |
| 20 | $10^{-6}$ | 0.2100    | 0.1964    |
| 20 | $10^{-6}$ | 0.2485    | 0.2202    |
| 20 | $10^{-6}$ | 0.2469    | 0.2485    |
| 20 | $10^{-6}$ | 0.2466    | 0.2485    |

Figure 4. Probability densities of $x_1$ estimation errors with the three filters in case 1.

Figure 5. Probability densities of $x_2$ estimation errors with the three filters in case 1.
6.2. Case 2

Consider a nonlinear system in which the state equation and the measurement equation are both nonlinear models:

\[
\begin{aligned}
    x_1(k+1) &= \cos\left(0.5x_1(k) + \frac{2.5x_2(k)}{1+x_2^2(k)+8\cos(1.2k)}\right) + w_1(k) \\
    x_2(k+1) &= \sin(x_1^2(k)) + w_2(k) \\
    y_1(k) &= \cos(x_1(k) + \sin(x_1^3(k))) + v_1(k+1) \\
    y_2(k) &= \sin(x_2(k) - \sin(x_2^3(k))) + v_2(k+1)
\end{aligned}
\]

where the initial value \(x(0)\) is a random value of \([0, 1]\), the initial estimation error covariance \(P(0|0) = 0.1 \times \text{diag}(1,1)\), and the process noise and measurement noise have the following characteristics:

\[
\begin{aligned}
    w_1(k) &\sim 0.9N(0,0.01) + 0.1N(0,0.2),
    w_2(k) &\sim 0.9N(0,0.02) + 0.1N(0,0.2) \\
    v_1(k) &\sim 0.9N(0,0.01) + 0.1N(0,0.2),
    v_2(k) &\sim 0.9N(0,0.02) + 0.1N(0,0.2)
\end{aligned}
\]

Figure 6 shows a diagram of the MTN identification system, while Figure 7 shows the estimated values of state variables \(x_1\) and \(x_2\) under the three filtering methods. Similar to case 1, the parameters are set at \(\varepsilon = 10^{-6}\). Tables 3 and 4 show the mean squared error and the mean relative error, respectively, of the estimated values under the three algorithms, which are computed as averages over 100 independent Monte Carlo runs, with each run containing 50 time steps. When \(\sigma = 5\), the three algorithms all obtain better filtering results. Figures 8 and 9 show the probability densities of the estimation errors when estimating the states \(x_1\) and \(x_2\), respectively, when the parameters are \(\varepsilon = 10^{-6}\) and \(\sigma = 5\). All of the results confirm that the proposed H-MCKF can outperform the MCKF significantly when the system is disturbed by non-Gaussian processes and measurement noise, and the H-MCKF_R further improves the filtering performance of the H-MCKF when the state and measurement equations are both nonlinear.

![Figure 6](image1.png)  
(a)  
![Figure 6](image2.png)  
(b)

Figure 6. (a) Graph of the MTN identify the state equation in case 2. (b) Graph of the MTN identify the measurement equation in case 2.
Figure 6. (a) Graph of the MTN identify the state equation in case 2. (b) Graph of the MTN identify the measurement equation in case 2.

Figure 7. (a,b) Set parameters: $\sigma = 5, \varepsilon = 10^{-6}$ in case 2; (c,d) set parameters: $\sigma = 10, \varepsilon = 10^{-6}$ in case 2.

Table 3. The mean squared error using the three methods in case 2.

| $\sigma$ | $\varepsilon$ | MSE of $x_1$ | MSE of $x_2$ |
|---------|-------------|---------------|---------------|
|         |             | MCEKF | H-MCKF | H-MCKF_R | MCEKF | H-MCKF | H-MCKF_R |
| $\sigma = 2$ | $\varepsilon = 10^{-6}$ | 0.4017 | 0.1230 | 0.1219 | 0.2090 | 0.0907 | 0.0883 |
| $\sigma = 5$ | $\varepsilon = 10^{-6}$ | 0.1148 | 0.1241 | 0.1233 | 0.2542 | 0.1200 | 0.1183 |
| $\sigma = 10$ | $\varepsilon = 10^{-6}$ | 0.3221 | 0.1254 | 0.1248 | 0.2220 | 0.1207 | 0.1193 |
| $\sigma = 20$ | $\varepsilon = 10^{-6}$ | 0.4040 | 0.1257 | 0.1251 | 0.2218 | 0.1208 | 0.1196 |

Table 4. The mean relative error using the three methods in case 2.

| $\sigma$ | $\varepsilon$ | MRE of $x_1$ | MRE of $x_2$ |
|---------|-------------|---------------|---------------|
|         |             | MCEKF | H-MCKF | H-MCKF_R | MCEKF | H-MCKF | H-MCKF_R |
| $\sigma = 2$ | $\varepsilon = 10^{-6}$ | 0.5106 | 0.2337 | 0.2306 | 0.3742 | 0.2355 | 0.2316 |
| $\sigma = 5$ | $\varepsilon = 10^{-6}$ | 0.2147 | 0.2551 | 0.2530 | 0.3070 | 0.2652 | 0.2645 |
| $\sigma = 10$ | $\varepsilon = 10^{-6}$ | 0.4527 | 0.2570 | 0.2553 | 0.3824 | 0.2661 | 0.2659 |
| $\sigma = 20$ | $\varepsilon = 10^{-6}$ | 0.4764 | 0.2575 | 0.2558 | 0.3789 | 0.2663 | 0.2662 |
The mean relative error using the three methods in case 2.

Figure 8. Probability densities of $x_1$ estimation errors with the three filters in case 2.

Table 4. The mean relative error using the three methods in case 2.

| Method          | $\varepsilon_1$ | $\varepsilon_2$ |
|-----------------|------------------|------------------|
| MCEKF H-MCKF    | 0.4764           | 0.4527           |
| H-MCKF R        | 0.2575           | 0.2570           |
| MCEKF H-MCKF R  | 0.2558           | 0.2553           |
| MCEKF H-MCKF    | 0.3789           | 0.3824           |
| H-MCKF R        | 0.2663           | 0.2661           |
| MCEKF H-MCKF    | 0.2662           | 0.2659           |

Figure 9. Probability densities of $x_2$ estimation errors with the three filters in case 2.

7. Conclusions

This paper considered a wide range of filter design problems for the state estimation of multivariable dynamic systems, which consist of a strong nonlinear dynamic model and a strong nonlinear observation model. Firstly, we transformed those strong nonlinear models into a higher order polynomial series using a multidimensional Taylor network. Secondly, all higher order items in the polynomial series were defined as hidden variables. Those higher order series were then rewritten as their pseudolinear equivalents. Thirdly, dynamic relationships between all hidden variables and known variables were constructed using the multidimensional Taylor network. Combining the original model of pseudolinearization with the higher order hidden variable dynamic model, linear dynamic models fitted to a standard Kalman filter were presented. Finally, considering that a finite number of samples from modeling error can be obtained, we built the higher order extended Kalman filter based on maximum correlation entropy, and acquired better filter performance than offered by the existing MCEKF [22].

Outlook: There exist several challenges worthy of further research. Firstly, the proposed higher order extended Kalman filter based on maximum correlation entropy is an online iteration process that obtains state estimation constantly, but, as such, it loses one important function possessed by the standard Kalman filter: the ability to operate in real time. Secondly, the linearized model parameters of the original nonlinear model and the hidden variable dynamic model were identified by local time period data; thus, they need to be updated with new time period data in order to fit the time dynamics of the system. Thirdly, in this paper, on the basis of defining all of the hidden variables, we
established a linear form of the strong nonlinear model in an expanded state with the original variables and all hidden variables, and obtained better estimation performance than that of a standard EKF; if measurements can be expanded in the same manner as state, we believe that such a filter may offer better estimation performance than the one established by this paper.

**Author Contributions:** Conceptualization, Q.W., X.S. and C.W.; methodology, C.W.; software, Q.W.; writing—original draft preparation, Q.W. writing—review and editing, Q.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the institutions: [01] key project of National Natural Science Foundation of China (No. 61751304), extraction and deep learning of optimal decision rules in an uncertain small sample environment; [02] key project of National Natural Science Foundation of China and Zhejiang joint fund for integration of industrialization and industrialization (No. u1509203), life cycle fault prediction, and intelligent health management of large ship power system operation; [03] key project of National Natural Science Foundation of China (No. 61933013): intelligent diagnosis, prediction and maintenance of abnormal working conditions of large petrochemical units.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Kalman, R.E. A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D J. Basic Eng.* 1960, 82, 35–45. [CrossRef]
2. Wen, C.; Wang, Z.; Liu, Q.; Alsaadi, F.E. Recursive distributed filtering for a class of state-saturated systems with fading measurements and quantization effects. *IEEE Trans. Syst. Man Cybern. Syst.* 2016, 48, 930–941. [CrossRef]
3. Wen, C.; Wang, Z.; Hu, J.; Liu, Q.; Alsaadi, F.E. Rlsaadi. Recursive filtering for state-saturated systems with randomly occurring nonlinearities and missing measurements. *Int. J. Robust Nonlinear Control.* 2018, 28, 1715–1727. [CrossRef]
4. Ge, Q.; Shao, T.; Duan, Z.; Wen, C. Performance Analysis of the Kalman Filter with Mismatched Noise Covariances. *IEEE Trans. Autom. Control.* 2016, 61, 4014–4019. [CrossRef]
5. Wen, C.; Cheng, X.; Xu, D.; Wen, C. Filter design based on characteristic functions for one class of multi-dimensional nonlinear non-Gaussian systems. *Automatica.* 2017, 82, 171–180. [CrossRef]
6. Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* 2002, 50, 174–188. [CrossRef]
7. Feng, X.; Wen, C.; Park, J.H. Sequential fusion $H_{\infty}$ filtering for multi-rate multi-sensor time-varying systems—A Krein-space approach. *IET Control Theory Appl.* 2017, 11, 369–381. [CrossRef]
8. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica.* 2017, 76, 70–77. [CrossRef]
9. Liu, X.; Qu, H.; Zhao, J.; Chen, B. Extended Kalman filter under maximum correntropy criterion. In Proceedings of the 2016 International Joint Conference on Neural Networks, Vancouver, BC, Canada, 24–29 July 2016; pp. 1733–1737.
10. Wang, G.; Li, N.; Zhang, Y. Maximum correntropy unscented Kalman and information filters for non-Gaussian measurement noise. *J. Frankl. Inst.* 2017, 354, 8659–8677. [CrossRef]
11. Meinholds, R.J.; Singpurwalla, N.D. Robustification of Kalman filter models. *J. Am. Stat. Assoc.* 1989, 84, 479–486. [CrossRef]
12. Wang, L.; Cheng, X.H.; Li, S.X. Gaussian Sum High Order Unscented Kalman Filtering Algorithm. *Chin. J. Electron.* 2017, 45, 424–430.
13. Zhang, C.; Yan, H.S. Identification of nonlinear time varying system with noise based on multi-dimensional Taylor network with optimal structure. *J. Southeast Univ.* 2017, 47, 1086–1093.
14. Wen, T.; Ge, Q.; Lyu, X.; Chen, L.; Constantinou, C.; Roberts, C.; Cai, B. A Cost-effective Wireless Network Migration Planning Method Supporting High-security Enabled Railway Data Communication Systems. *J. Frankl. Inst.* 2021, 358, 131–150. [CrossRef]
15. Wen, T.; Wen, C.; Roberts, C.; Cai, B. Distributed Filtering for a Class of Discrete-time Systems Over Wireless Sensor Networks. *J. Frankl. Inst.* 2020, 357, 3038–3055. [CrossRef]
16. Xiaohui, S.; Chenglin, W.; Tao, W. A Novel Step-by-Step High-Order Extended Kalman Filter Design for a Class of Complex Systems with Multiple Basic Multipliers. *Chin. J. Electron.* 2021, 30, 313–321. [CrossRef]
17. Xiaohui, S.; Chenglin, W.; Tao, W. High-Order Extended Kalman Filter Design for a Class of Complex Dynamic Systems with Polynomial Nonlinearities. *Chin. J. Electron.* 2021, 30, 508–515. [CrossRef]
18. Feng, X.; You, B. Random attractors for the two-dimensional stochastic g-Navier-Stokes equations. *Stochastics* 2019, 92, 613–626. [CrossRef]
19. Liu, W.; Chi, Y.; Zhang, G. Multiple Resolvable Group Estimation Based on the GLMB Filter with Graph Structure. In Proceedings of the 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Tianjin, China, 19–23 July 2018; pp. 960–964.
20. Wu, Z.; Shi, J.; Zhang, X.; Ma, W.; Chen, B. Kernel recursive maximum correntropy. *Signal Process.* 2015, 117, 11–16. [CrossRef]
21. Anderson, B.; Moore, J. *Optimal Filtering*; Prentice-Hall: New York, NY, USA, 1979.

22. Julier, S.; Uhlmann, J.; Durrant-Whyte, H.F. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Control.* **2000**, *45*, 477–482. [CrossRef]