Fractional central difference Kalman filter with unknown prior information

Tianyu Liu, Songsong Cheng, Yiheng Wei, Ang Li, Yong Wang*

Department of Automation, University of Science and Technology of China, Hefei, 230027, China

Abstract

In this paper, a generalized fractional central difference Kalman filter for nonlinear discrete fractional dynamic systems is proposed. Based on the Stirling interpolation formula, the presented algorithm can be implemented as no derivatives are needed. Besides, in order to estimate the state with unknown prior information, a maximum a posteriori principle based adaptive fractional central difference Kalman filter is derived. The adaptive algorithm can estimate the noise statistics and system state simultaneously. The unbiasedness of the proposed algorithm is analyzed. Several numerical examples demonstrate the accuracy and effectiveness of the two Kalman filters.

Keywords: Fractional calculus, Adaptive filter, Fractional Kalman filter, Central difference Kalman filter, Maximum a posteriori principle.

1. Introduction

The optimum Kalman filter is a recursive state estimation algorithm for integer order linear state space systems. It is widely used in numerous en-

*Corresponding author. E-mail: yongwang@ustc.edu.cn

gineering applications, such as aerospace, navigation [1], econometrics, computer vision [2], autopilots [3] and many others where estimation is relevant. The accuracy of the Kalman filter depends largely on certain assumptions, such as noise statistics. The problem is observed that the noise prior knowledge is unknown or time-varying in circumstances. The adaptive Kalman filter is a common tool to deal with this problem.

The classical Kalman filter was applied to the estimation problem for discrete dynamic systems [4]. Then based on the Taylor series approximation, Bucy and Sunahara put forward the extended Kalman filter (EKF) [5, 6]. Although the EKF is widely used for various engineering fields, there still exist some theoretical limitations, fox example, nonlinear functions must be continuously differentiable and the filter is required to calculate the Jacobian matrix. Following the intuition that "*it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation*", using the unscented transformation, Julier and Uhlmann et al. presented a new approach to approximate the posterior mean and the posterior error covariance [7]. The corresponding filter is known as the unscented Kalman filter (UKF). The UKF ensures an accuracy of at least the second order Taylor series approximation. But the implementation of a UKF is more computationally expensive than an EKF. Therefore, Biswas et al. proposed a new single propagated unscented Kalman filter and an extrapolated single propagated unscented Kalman filter to reduce computational complexity [8].

For nonlinear Gaussian systems, Ito et al. presented the systematic formulation of Gaussian optimal recursive filters, and obtained a novel central difference filter [9]. At the same time, NøRgaard et al. utilized the Stirling interpolation formula to approximate the posterior mean and the posterior covariance. Then the divided difference filter is developed [10]. Those two filters are essentially identical and can be referred to as the central difference Kalman filter (CDKF) [11].

The performance of the KF depends largely on prior information of noise statistics. The use of imprecise information will result in estimation errors or even filtering divergence. Adaptive filtering is an effective way to solve this problem. Most of the adaptive filtering methods are applied to linear systems. It can be divided into four categories: Bayesian, maximum likelihood, correlation and covariance matching [12]. Based on the maximum a posteriori (MAP) principle, the popular Sage-Husa AKF (SHAKF) [13], which estimates the noise statistics and state recursively , also can be considered as a covariance matching method. Besides, the variational Bayesian based AKF is also an approximation of the Bayesian method [14]. For nonlinear systems, several approaches are investigated.

On the other hand, thanks to that many systems can be described accurately with the introduction of fractional calculus, fractional systems have attracted much attention from engineering and physics fields. Besides, the application of fractional calculus in control systems also has rapidly development, especially in stability analysis [15, 16], controller design [17, 18], adaptive filtering [19, 20], etc. An important class of theoretical and practical problems is how to obtain the exact state when state variables cannot be measured directly. Motivated by this, the fractional Kalman filter (FKF) and the fractional extended Kalman filter (FEKF) are proposed [21]. The

FKF algorithm is used for state estimation in the systems with ultracapacitor [22], fractional nonlinear systems in a chaotic communication scheme [23] and over networks with packet losses [24], etc. The prime difference between the FKF and the integer Kalman filter is that the integer order dynamic systems can be considered as a Markov process, but fractional dynamic systems can not. Because of the existence of the fractional differential operator, the estimated state $\mathbf{x}_t$ of the FKF depends on all of the previous state, which leads to significant complexity. Meanwhile the defects of the integer order EKF also exist in the FEKF.

Motivated by the previous discussions, a generalized fractional central difference Kalman filter (FCDKF) is presented. Based on the conventional CDKF, the proposed FCDKF is also a derivative-free filtering algorithm. Furthermore, considering that the prior information is hard to obtain, an adaptive fractional central difference Kalman filter (AFCDKF) is addressed, which can evaluate the system state and noise statistics simultaneously. The main contributions are concluded as follows

- A FCDKF and an AFCDKF are addressed to estimate the system state for different prior information conditions;
- The unbiasedness of the AFCDKF algorithm is analyzed and then an unbiased recursive algorithm is developed;
- The approximate accuracy and numerical complexity of proposed algorithms are analyzed.

The rest of this paper is organized as follows. Section 2 reviews the fundamental knowledge of fractional calculus and CDKF. The FCDKF and

AFCDKF for fractional discrete nonlinear systems with stochastic perturbation are designed in Section 3. Section 4 provides several illustrative numerical examples. Finally, Section 5 draws some conclusions.

2. Preliminaries

2.1. Problem statement

The fractional discrete nonlinear system with stochastic perturbation can be described as follow

Definition 2.1 *The fractional discrete nonlinear system with stochastic perturbation can be described as*

$$\begin{cases} \nabla^{\boldsymbol{\alpha}} \mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\omega}_{k-1}, \\ \quad \mathbf{x}_k = \nabla^{\boldsymbol{\alpha}} \mathbf{x}_k - \sum_{j=1}^{k} (-1)^j \boldsymbol{\gamma}_j \mathbf{x}_{k-j}, \\ \quad \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\nu}_k, \end{cases} \tag{1}$$

where $\nabla^{\boldsymbol{\alpha}} = [\nabla^{\alpha_1}, \cdots, \nabla^{\alpha_n}]^{\mathrm{T}}$ *and* $\boldsymbol{\gamma}_j = \mathrm{diag}\left[\binom{\alpha_1}{j}, \cdots, \binom{\alpha_n}{j}\right]$.

Here k denotes the time index, $\mathbf{x}_k \in \mathbb{R}^n$, $\boldsymbol{\alpha} \in \mathbb{R}^n$, and $\mathbf{z}_k \in \mathbb{R}^m$ are the system state, orders of difference and measurement value, respectively. $\mathbf{f}_k : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathbf{h}_k : \mathbb{R}^n \to \mathbb{R}^m$ are the nonlinear state transform function and measurement function. $\boldsymbol{\omega}_k \in \mathbb{R}^n$ and $\boldsymbol{\nu}_k \in \mathbb{R}^m$ mean the system noise and measurement noise. Moreover, $\hat{\mathbf{x}}_{i|j} = \mathrm{E}\{\mathbf{x}_i \mid \mathbf{Z}_j\}$ indicates the state mean conditioned on $\mathbf{Z}_j$, where $\mathbf{Z}_j = [\mathbf{z}_1, \cdots, \mathbf{z}_j]$ is the the observed value. ∇ is the nabla operator, and its definition is given by Definition 2.2.

Definition 2.2 *The fractional backward difference of the order* α *is given by*

$$\nabla^{\alpha} f(k) = \sum_{j=0}^{k} (-1)^j \binom{\alpha}{j} f(k-j), \tag{2}$$

where $k \in \mathbb{N}_+$ and the corresponding binomial coefficient can be defined as $\binom{\alpha}{j} = \frac{\alpha(\alpha-1)\ \cdots\ (\alpha-j+1)}{j!}$.

The same as the integer order EKF, the FEKF has been proposed to estimate the system state. But the Jacobian matrix of nonlinear functions is also required in FEKF, which is one of the major constraints. Furthermore, the performance of state estimation is positively related to the accuracy of prior noise information. In most situations, those statistics are inexactly known or even completely unknown. This will lead to large estimation errors or even to filtering divergence. Therefore, the objective of this paper is to design a derivative-free FKF algorithm to estimate the system state exactly. In addition, the adaptive FKF with unknown prior information is also investigated, which aims to evaluate the system state and noise statistics concurrently.

To simplify the analysis, the following common assumptions are carried out [25].

Assumption 2.3 *The two noise vectors subject to Gaussian distribution*

$$\begin{cases} \mathrm{E}\{\boldsymbol{\omega}_k\} = \mathbf{q}_k,\ \mathrm{Cov}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) = \mathbf{Q}_i\delta_{ij}, \\ \mathrm{E}\{\boldsymbol{\nu}_k\} = \mathbf{r}_k,\ \mathrm{Cov}(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j) = \mathbf{R}_i\delta_{ij}, \quad \forall\ i, j, k, \\ \mathrm{Cov}(\boldsymbol{\omega}_i, \boldsymbol{\nu}_j) = \mathbf{0}, \end{cases} \tag{3}$$

where δ_{ij} is the Kronecher-δ function, $\mathbf{R}$ is a positive definite matrix and $\mathbf{Q}$ is a positive semidefinite matrix.

Assumption 2.4 *The initial state $\mathbf{x}_0$ obeys Gaussian distribution, and is uncorrelated with both the system and measurement noises.*

Assumption 2.5 $\mathrm{E}\{\mathbf{x}_i \mid \mathbf{Z}_j\} = \mathrm{E}\{\mathbf{x}_i \mid \mathbf{Z}_i\} = \hat{\mathbf{x}}_i,\ \forall\ i \leq j$.

Assumption 2.6 $\mathrm{E}\{(\mathbf{x}_i - \hat{\mathbf{x}}_i)(\mathbf{x}_j - \hat{\mathbf{x}}_j)^{\mathrm{T}}\} = \mathbf{0},\ \forall\ i \neq j.$

2.2. Fundamental knowledge

First, the Stirling interpolation formula is introduced.

Definition 2.7 *Assuming that* $\mathbf{x} \in \mathbb{R}^n$*,* $\mathbf{z} = \mathbf{f}(\mathbf{x})$ *is a multidimensional differentiable function, applying the Stirling interpolation formula around the point* $\mathbf{x} = \bar{\mathbf{x}}$ *yields*

$$\mathbf{z} = \mathbf{f}(\mathbf{x}) = \mathbf{f}(\bar{\mathbf{x}} + \Delta\mathbf{x}) = \mathbf{f}(\bar{\mathbf{x}}) + \widetilde{\mathbf{D}}_{\Delta\mathbf{x}}\mathbf{f} + \cdots, \tag{4}$$

where $\widetilde{\mathbf{D}}_{\Delta\mathbf{x}}\mathbf{f} = \frac{1}{\hbar}\big(\sum_{i=1}^{n} \Delta x_i \mu_i \delta_i\big)\mathbf{f}(\bar{\mathbf{x}})$ *and* $\Delta\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}}$ *.*

Here, $\hbar$ denotes a selected interval length, and μ_i and δ_i are the locally difference operators (see [10]).

Next, the so-called Cholesky factorization is introduced. Considering the function $\mathbf{z} = \mathbf{f}(\mathbf{x})$, the stochastic state $\mathbf{x}$ takes on a Gaussian distribution, denoted as $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$. Based on the Stirling interpolation formula, the probability distribution of $\mathbf{z} \sim \mathcal{N}(\bar{\mathbf{z}}, \mathbf{P}_{\mathbf{z}})$ can be deduced. Based on the Cholesky factorization, we derive $\mathbf{P}_{\mathbf{x}} = \mathbf{S}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{\mathrm{T}}$. Next, the following transformation of $\mathbf{x}$ is introduced:

$$\mathbf{y} = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{x}, \tag{5}$$

$$\widetilde{\mathbf{f}}(\bar{\mathbf{y}}) = \mathbf{f}(\mathbf{S}_{\mathbf{x}}\bar{\mathbf{y}}) = \mathbf{f}(\bar{\mathbf{x}}). \tag{6}$$

The following results can be derived [10]

$$\bar{\mathbf{y}} = \mathrm{E}\{\mathbf{y}\} = \mathbf{S}_{\mathbf{x}}^{-1}\bar{\mathbf{x}}, \tag{7}$$

$$\mathrm{E}\{(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^{\mathrm{T}}\} = \mathbf{I}, \tag{8}$$

$$\mathrm{E}\{(\mathbf{y}_i-\bar{\mathbf{y}}_i)(\mathbf{y}_j-\bar{\mathbf{y}}_j)^{\mathrm{T}}\}=\mathbf{0},\ \forall\ i\neq j. \tag{9}$$

Using the linear transformation, each element of stochastic state $\mathbf{y}$ is independent. Similarly, each element of $\Delta\mathbf{y}$ is also irrelevant.

3. Main Results

3.1. Fractional central difference Kalman filter

In this section, the proposed FCDKF algorithm is deduced. Furthermore, the approximate accuracy and numerical complexity are analyzed briefly.

3.1.1. Implementation of FCDKF

To simplify, the nonlinear function will be replaced by a first-order Stirling interpolation approximation (Def 2.7). The estimated state $\hat{\mathbf{x}}_{k-1}$ and covariance $\mathbf{P}_{k-1}$ are known.

Firstly, the predicted state $\hat{\mathbf{x}}_{k|k-1}$ is given by

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1}&=\mathrm{E}\{\mathbf{x}_k\mid\mathbf{Z}_{k-1}\}\\&=\mathrm{E}\{\mathbf{f}(\mathbf{x}_{k-1})+\boldsymbol{\omega}_{k-1}-\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{x}_{k-j}\mid\mathbf{Z}_{k-1}\}\\&\approx\mathrm{E}\{[\mathbf{f}(\hat{\mathbf{x}}_{k-1})+\widetilde{\mathbf{D}}_{\Delta\mathbf{x}_{k-1}}\mathbf{f}]\mid\mathbf{Z}_{k-1}\}+\mathrm{E}\{\boldsymbol{\omega}_{k-1}\mid\mathbf{Z}_{k-1}\}\\&\quad-\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathrm{E}\{\mathbf{x}_{k-j}\mid\mathbf{Z}_{k-1}\}.\end{aligned}\tag{10}$$

Considering Assumption 2.3 and Assumption 2.5, it can be obtained that $\mathrm{E}\{\boldsymbol{\omega}_{k-1}\mid\mathbf{Z}_{k-1}\}=\mathbf{q}$ and $\mathrm{E}\{\mathbf{x}_{k-j}\mid\mathbf{Z}_{k-1}\}=\hat{\mathbf{x}}_{k-j}$. Then, substituting (6) into (10) yields

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1}&\approx\mathrm{E}\{[\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1})+\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}}]\mid\mathbf{Z}_{k-1}\}-\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j}+\mathbf{q}\\&=\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1})+\mathrm{E}\{\tfrac{1}{\hbar}(\textstyle\sum_{i=1}^{n}\Delta y_{k-1}^{i}\mu_i\delta_i)\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1})\mid\mathbf{Z}_{k-1}\}\\&\quad-\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j}+\mathbf{q}.\end{aligned}\tag{11}$$

Because $\Delta\mathbf{y}_{k-1} = \mathbf{y}_{k-1} - \hat{\mathbf{y}}_{k-1} = \mathbf{S}_{k-1}^{-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) = \mathbf{S}_{k-1}^{-1}\Delta\mathbf{x}_{k-1}$, one has $\mathrm{E}\{\Delta\mathbf{y}_{k-1} \mid \mathbf{Z}_{k-1}\} = \mathbf{S}_{k-1}^{-1}\mathrm{E}\{\Delta\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1}\} = \mathbf{0}$. So the predicted state can be described by

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &\approx \widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1}) + \tfrac{1}{\hbar}\textstyle\sum_{i=1}^{n}\mu_i\delta_i\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1})\mathrm{E}\{\Delta y_{k-1}^{i} \mid \mathbf{Z}_{k-1}\} \\ &\quad - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j} + \mathbf{q} \\ &= \mathbf{f}(\hat{\mathbf{x}}_{k-1}) - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j} + \mathbf{q}.\end{aligned} \tag{12}$$

Here, $\mathbf{S}_{k-1} = [\mathbf{s}_{k-1}^1, \mathbf{s}_{k-1}^2, \cdots, \mathbf{s}_{k-1}^n]$ represents the Cholesky factor, which can be obtained by $\mathbf{P}_{k-1} = \mathbf{S}_{k-1}\mathbf{S}_{k-1}^{\mathrm{T}}$.

Next, the prediction error covariance $\mathbf{P}_{k|k-1}$ can be formulated as

$$\mathbf{P}_{k|k-1} = \mathrm{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}\}. \tag{13}$$

Using (11) yields

$$\begin{aligned}\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}(\mathbf{x}_{k-1}) + \boldsymbol{\omega}_{k-1} - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{x}_{k-j} - \mathbf{f}(\hat{\mathbf{x}}_{k-1}) \\ &\quad + \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j} - \mathbf{q} \\ &= \widetilde{\mathbf{D}}_{\Delta\mathbf{x}_{k-1}}\mathbf{f} - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\Delta\mathbf{x}_{k-j} + \boldsymbol{\omega}_{k-1} - \mathbf{q}.\end{aligned} \tag{14}$$

Substituting (14) into (13) results in

$$\begin{aligned}\mathbf{P}_{k|k-1} &= \mathrm{E}\{[\widetilde{\mathbf{D}}_{\Delta\mathbf{x}_{k-1}}\mathbf{f} + \boldsymbol{\omega}_{k-1} - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\Delta\mathbf{x}_{k-j} - \mathbf{q}] \\ &\quad \times[\widetilde{\mathbf{D}}_{\Delta\mathbf{x}_{k-1}}\mathbf{f} + \boldsymbol{\omega}_{k-1} - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\Delta\mathbf{x}_{k-j} - \mathbf{q}]^{\mathrm{T}}\} \\ &= \mathrm{E}\{[\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\mathbf{f} + \boldsymbol{\omega}_{k-1} - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j} - \mathbf{q}] \\ &\quad \times[\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\mathbf{f} + \boldsymbol{\omega}_{k-1} - \textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j} - \mathbf{q}]^{\mathrm{T}}\} \\ &= \mathrm{E}\{[\textstyle\sum_{i=1}^{k}(-1)^i\boldsymbol{\gamma}_i\mathbf{S}_{k-i}\Delta\mathbf{y}_{k-i}][\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}]^{\mathrm{T}}\} \\ &\quad -\mathrm{E}\{\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}}[\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}]^{\mathrm{T}}\} \\ &\quad -\mathrm{E}\{[\textstyle\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}](\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}})^{\mathrm{T}}\} \\ &\quad +\mathrm{E}\{\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}}(\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}})^{\mathrm{T}}\} + \mathbf{Q}.\end{aligned} \tag{15}$$

The fourth term in (15) has already been resolved in [10]. Here, the result is given directly as follow

$$\begin{aligned}&\mathrm{E}\{\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}}(\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}})^{\mathrm{T}}\}\\&=\tfrac{1}{4\hbar^2}\textstyle\sum_{n}^{i=1}[\mathbf{f}(\hat{\mathbf{x}}_{k-1}+\hbar\mathbf{s}_{k-1}^{i})-\mathbf{f}(\hat{\mathbf{x}}_{k-1}-\hbar\mathbf{s}_{k-1}^{i})]\\&\quad\times[\mathbf{f}(\hat{\mathbf{x}}_{k-1}+\hbar\mathbf{s}_{k-1}^{i})-\mathbf{f}(\hat{\mathbf{x}}_{k-1}-\hbar\mathbf{s}_{k-1}^{i})]^{\mathrm{T}}.\end{aligned}\tag{16}$$

Then for the first term, we have

$$\begin{aligned}&\mathrm{E}\{[\textstyle\sum_{i=1}^{k}(-1)^{i}\boldsymbol{\gamma}_i\mathbf{S}_{k-i}\Delta\mathbf{y}_{k-i}][\sum_{j=1}^{k}(-1)^{j}\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}]^{\mathrm{T}}\}\\&=\textstyle\sum_{i=1,j=i}^{k}\boldsymbol{\gamma}_i\mathbf{S}_{k-i}\mathrm{E}\{\Delta\mathbf{y}_{k-i}(\Delta\mathbf{y}_{k-j})^{\mathrm{T}}\}\mathbf{S}_{k-j}^{\mathrm{T}}\boldsymbol{\gamma}_j^{\mathrm{T}}\\&\quad+\textstyle\sum_{i=1}^{k}\sum_{j=1,i\neq j}^{k}\boldsymbol{\gamma}_i\mathbf{S}_{k-i}\mathrm{E}\{\Delta\mathbf{y}_{k-i}(\Delta\mathbf{y}_{k-j})^{\mathrm{T}}\}\mathbf{S}_{k-j}^{\mathrm{T}}\boldsymbol{\gamma}_j^{\mathrm{T}}.\end{aligned}\tag{17}$$

Employing (8) and (9) into (17) yields

$$\begin{aligned}&\mathrm{E}\{[\textstyle\sum_{i=1}^{k}(-1)^{i}\boldsymbol{\gamma}_i\mathbf{S}_{k-i}\Delta\mathbf{y}_{k-i}][\sum_{j=1}^{k}(-1)^{j}\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}]^{\mathrm{T}}\}\\&=\textstyle\sum_{j=1}^{k}\boldsymbol{\gamma}_j\mathbf{P}_{k-j}\boldsymbol{\gamma}_j^{\mathrm{T}}.\end{aligned}\tag{18}$$

For the second and third terms, defining $\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1}=[\mu_1\delta_1\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1}),\mu_2\delta_2\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1}),\cdots,\mu_n\delta_n\widetilde{\mathbf{f}}(\hat{\mathbf{y}}_{k-1})]$ and $\Delta\mathbf{y}_{k-1}=\mathbf{y}_{k-1}-\hat{\mathbf{y}}_{k-1}=[\Delta y_{k-1}^{1},\Delta y_{k-1}^{2},\cdots,\Delta y_{k-1}^{n}]^{\mathrm{T}}$, so $\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}}=\frac{1}{\hbar}\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1}\Delta\mathbf{y}_{k-1}$, one has

$$\begin{aligned}&\mathrm{E}\{[\textstyle\sum_{j=1}^{k}(-1)^{j}\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}](\widetilde{\mathbf{D}}_{\Delta\mathbf{y}_{k-1}}\widetilde{\mathbf{f}})^{\mathrm{T}}\}\\&=\mathrm{E}\{[\textstyle\sum_{j=1}^{k}(-1)^{j}\boldsymbol{\gamma}_j\mathbf{S}_{k-j}\Delta\mathbf{y}_{k-j}](\frac{1}{\hbar}\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1}\Delta\mathbf{y}_{k-1})^{\mathrm{T}}\}\\&=-\tfrac{1}{\hbar}\boldsymbol{\gamma}_1\mathbf{S}_{k-1}(\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1})^{\mathrm{T}}.\end{aligned}\tag{19}$$

In total, (13) can be reformulated as

$$
\begin{aligned}
\mathbf{P}_{k|k-1} &= \mathrm{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}\} \\
&= \mathrm{E}\{\widetilde{\mathbf{D}}_{\Delta \mathbf{y}_{k-1}} \widetilde{\mathbf{f}} (\widetilde{\mathbf{D}}_{\Delta \mathbf{y}_{k-1}} \widetilde{\mathbf{f}})^{\mathrm{T}}\} - \tfrac{1}{\hbar} \mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1} \mathbf{S}_{k-1}^{\mathrm{T}} \boldsymbol{\gamma}_1^{\mathrm{T}} \\
&\quad - \tfrac{1}{\hbar} \boldsymbol{\gamma}_1 \mathbf{S}_{k-1} (\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1})^{\mathrm{T}} + \textstyle\sum_{j=1}^{k} \boldsymbol{\gamma}_j \mathbf{P}_{k-j} \boldsymbol{\gamma}_j^{\mathrm{T}} + \mathbf{Q} \\
&= \tfrac{1}{4\hbar^2} \textstyle\sum_{i=1}^{n} [\mathbf{f}(\hat{\mathbf{x}}_{k-1} + \hbar \mathbf{s}_{k-1}^{i}) - \mathbf{f}(\hat{\mathbf{x}}_{k-1} - \hbar \mathbf{s}_{k-1}^{i})] \\
&\quad \times [\mathbf{f}(\hat{\mathbf{x}}_{k-1} + \hbar \mathbf{s}_{k-1}^{i}) - \mathbf{f}(\hat{\mathbf{x}}_{k-1} - \hbar \mathbf{s}_{k-1}^{i})]^{\mathrm{T}} - \tfrac{1}{\hbar} \mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1} \mathbf{S}_{k-1}^{\mathrm{T}} \boldsymbol{\gamma}_1^{\mathrm{T}} \\
&\quad - \tfrac{1}{\hbar} \boldsymbol{\gamma}_1 \mathbf{S}_{k-1} (\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1})^{\mathrm{T}} + \textstyle\sum_{j=1}^{k} \boldsymbol{\gamma}_j \mathbf{P}_{k-j} \boldsymbol{\gamma}_j^{\mathrm{T}} + \mathbf{Q}.
\end{aligned}
\tag{20}
$$

The deduction of measurement update is similar to the integer order central difference Kalman filter. The output prediction is given by

$$
\begin{aligned}
\hat{\mathbf{z}}_{k|k-1} &= \mathrm{E}\{\mathbf{h}(\mathbf{x}_k) + \boldsymbol{\nu}_k \mid \mathbf{Z}_{k-1}\} \\
&= \mathrm{E}\{\widetilde{\mathbf{h}}(\hat{\mathbf{y}}_{k|k-1}) + \widetilde{\mathbf{D}}_{\Delta \mathbf{y}_{k|k-1}} \widetilde{\mathbf{h}} + \boldsymbol{\nu}_k \mid \mathbf{Z}_{k-1}\} \\
&= \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{r},
\end{aligned}
\tag{21}
$$

and the covariance

$$
\begin{aligned}
\mathbf{P}_{\widetilde{z}_k} &= \mathrm{E}\{(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\widetilde{\mathbf{D}}_{\Delta \mathbf{x}_{k|k-1}} \mathbf{h} + \boldsymbol{\nu}_k - \mathbf{r})(\widetilde{\mathbf{D}}_{\Delta \mathbf{x}_{k|k-1}} \mathbf{h} + \boldsymbol{\nu}_k - \mathbf{r})^{\mathrm{T}}\} \\
&= \tfrac{1}{4\hbar^2} \textstyle\sum_{i=1}^{n} [\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar \mathbf{s}_{k|k-1}^{i}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar \mathbf{s}_{k|k-1}^{i})] \\
&\quad \times [\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar \mathbf{s}_{k|k-1}^{i}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar \mathbf{s}_{k|k-1}^{i})]^{\mathrm{T}} + \mathbf{R}.
\end{aligned}
\tag{22}
$$

According to (14), the predicted error cross-covariance can be written as

$$
\begin{aligned}
\mathbf{P}_{\widetilde{x}_k \widetilde{z}_k} &= \mathrm{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathbf{S}_{k|k-1} \Delta \mathbf{y}_{k|k-1})(\widetilde{\mathbf{D}}_{\Delta \mathbf{y}_{k|k-1}} \widetilde{\mathbf{h}} + \boldsymbol{\nu}_k - \mathbf{r})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathbf{S}_{k|k-1} \Delta \mathbf{y}_{k|k-1})[\tfrac{1}{\hbar} (\textstyle\sum_{i=1}^{n} \Delta y_{k|k-1}^{i} \mu_i \delta_i) \widetilde{\mathbf{h}}(\hat{\mathbf{y}}_{k|k-1})]^{\mathrm{T}}\} \\
&= \tfrac{1}{2\hbar} \textstyle\sum_{i=1}^{n} \left\{ \mathbf{s}_{k|k-1}^{i} [\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar \mathbf{s}_{k|k-1}^{i}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar \mathbf{s}_{k|k-1}^{i})]^{\mathrm{T}} \right\},
\end{aligned}
\tag{23}
$$

where the Cholesky factor $\mathbf{S}_{k|k-1} = [\mathbf{s}_{k|k-1}^{1}, \mathbf{s}_{k|k-1}^{2}, \cdots, \mathbf{s}_{k|k-1}^{n}]$ is derived by $\mathbf{P}_{k|k-1} = \mathbf{S}_{k|k-1} \mathbf{S}_{k|k-1}^{\mathrm{T}}$.

Then, the estimated state $\hat{\mathbf{x}}_k$, Kalman filtering gain $\mathbf{K}_k$, and the state estimation covariance $\mathbf{P}_k$ can be deduced.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}), \tag{24}$$

$$\mathbf{K}_k = \mathbf{P}_{\widetilde{\mathbf{x}}_k \widetilde{\mathbf{z}}_k} \mathbf{P}_{\widetilde{\mathbf{z}}_k}^{-1}, \tag{25}$$

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{\widetilde{\mathbf{z}}_k} \mathbf{K}_k^{\mathrm{T}}. \tag{26}$$

Combining time updating (11), (20) and measurement updating (21)–(26), the proposed FCDKF operates recursively, whose pseudocode is shown in Algorithm 1.

Algorithm 1 Fractional central difference Kalman filter

Initialization:

Set the system initial values: $\mathbf{x}_0, \mathbf{P}_0$

Set the noise stochastic values: $\mathbf{q}, \mathbf{r}, \mathbf{Q}, \mathbf{R}$

Set the short memory principle length: L

Set the interval length: $\hbar$

On-line updating:

for $k = 1 \to K$ **do**

Cholesky decomposition:

$$\mathbf{P}_{k-1} = \mathbf{S}_{k-1}\mathbf{S}_{k-1}^{\mathrm{T}} \qquad \mathbf{S}_{k-1} = [\mathbf{s}_{k-1}^1, \mathbf{s}_{k-1}^2, \cdots, \mathbf{s}_{k-1}^n]$$

time updating:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1}) - \sum_{j=k-L+1}^{k} (-1)^j \boldsymbol{\gamma}_j \hat{\mathbf{x}}_{k-j} + \mathbf{q} \qquad \triangleright \text{ state prediction}$$

$$\begin{aligned}\mathbf{P}_{k|k-1} = \tfrac{1}{4\hbar^2} \sum_{i=1}^{n} &[\mathbf{f}(\hat{\mathbf{x}}_{k-1} + \hbar \mathbf{s}_{k-1}^i) - \mathbf{f}(\hat{\mathbf{x}}_{k-1} - \hbar \mathbf{s}_{k-1}^i)][\mathbf{f}(\hat{\mathbf{x}}_{k-1} + \hbar \mathbf{s}_{k-1}^i) \\ &- \mathbf{f}(\hat{\mathbf{x}}_{k-1} - \hbar \mathbf{s}_{k-1}^i)]^{\mathrm{T}} + \mathbf{Q} - \tfrac{1}{\hbar} \mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1} \mathbf{S}_{k-1}^{\mathrm{T}} \boldsymbol{\gamma}_1^{\mathrm{T}} - \tfrac{1}{\hbar} \boldsymbol{\gamma}_1 \mathbf{S}_{k-1} (\mathbf{G}_{\widetilde{\mathbf{f}}}^{k-1})^{\mathrm{T}} \\ &+ \sum_{j=k-L+1}^{k} \boldsymbol{\gamma}_j \mathbf{P}_{k-j} \boldsymbol{\gamma}_j^{\mathrm{T}}\end{aligned}$$

$\triangleright$ state prediction error covariance

Cholesky decomposition:

$$\mathbf{P}_{k|k-1} = \mathbf{S}_{k|k-1}\mathbf{S}_{k|k-1}^{\mathrm{T}} \qquad \mathbf{S}_{k|k-1} = [\mathbf{s}_{k|k-1}^{1}, \mathbf{s}_{k|k-1}^{2}, \cdots, \mathbf{s}_{k|k-1}^{n}]$$

measurement updating:

$$\mathbf{z}_{k|k-1} = \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{r} \qquad \triangleright \text{ measurement value estimation}$$

$$\mathbf{P}_{\widetilde{z}_k} = \frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}_{k|k-1}^{i}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar\mathbf{s}_{k|k-1}^{i})] \times [\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}_{k|k-1}^{i}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar\mathbf{s}_{k|k-1}^{i})]^{\mathrm{T}} + \mathbf{R}$$

$\triangleright$ measurement prediction error covariance

$$\mathbf{P}_{\widetilde{x}_k\widetilde{z}_k} = \frac{1}{2\hbar}\sum_{i=1}^{n}\left\{\mathbf{s}_{k|k-1}^{i}[\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}_{k|k-1}^{i}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar\mathbf{s}_{k|k-1}^{i})]^{\mathrm{T}}\right\}$$

$\triangleright$ prediction error cross-covariance

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \qquad \triangleright \text{ state estimation}$$

$$\mathbf{K}_k = \mathbf{P}_{\widetilde{x}_k\widetilde{z}_k}\mathbf{P}_{\widetilde{z}_k}^{-1} \qquad \triangleright \text{ Kalman gain}$$

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{P}_{\widetilde{z}_k}\mathbf{K}_k^{\mathrm{T}} \qquad \triangleright \text{ state estimation error covariance}$$

end for

Remark 3.1 *As analyzed in [10], high order error terms between the Stirling interpolation formula and the Taylor series formula are controlled by* $\hbar$. *A reasonable choice of* $\hbar$ *makes the Stirling interpolation more attractive than the Taylor series. The selection of* $\hbar$ *depends on the approximated function.*

Remark 3.2 *In order to ensure that the measurement prediction error covariance* $\mathbf{P}_{\widetilde{z}_k}$ *is invertible, the case of* $\mathbf{R}_k$ *being positive definite is required. Actually, the case that* $\mathbf{R}_k$ *is positive definite is a sufficient condition for the statement that* $\mathbf{P}_{\widetilde{z}_k}$ *is invertible. Nevertheless, this condition does encompass the vast majority of applications of practical interest [25], so* $\mathbf{R}_k$ *being positive definite is a common and standard assumption in most literatures as well as in this paper.*

3.1.2. Performance analysis

First, the approximate accuracy is analyzed briefly. For convenience, the fractional dynamic system (Def. 2.1) can be converted into

$$\begin{cases} \mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) - \sum_{j=1}^{k}(-1)^j \boldsymbol{\gamma}_j \mathbf{x}_{k-j} + \boldsymbol{\omega}_{k-1}, \\ \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\nu}_k. \end{cases} \tag{27}$$

The algorithm performance is mainly influenced by the short memory length L and the approximate accuracy of nonlinear functions $\mathbf{f}(\mathbf{x}_k)$ and $\mathbf{h}(\mathbf{x}_k)$.

In [10], based on the Taylor series expansion, the second order Stirling approximate accuracy for an arbitrary function $z = f(x)$ is addressed. The first Stirling approximation is given

$$\begin{aligned} f(\bar{x}) \approx f(\bar{x}) + f^{'}(\bar{x})(x-\bar{x}) + \tfrac{\hbar^2 f^{(3)}(\bar{x})}{3!}(x-\bar{x}) \\ + \tfrac{\hbar^4 f^{(5)}(\bar{x})}{5!}(x-\bar{x}) + \tfrac{\hbar^6 f^{(7)}(\bar{x})}{7!}(x-\bar{x}) + \ldots. \end{aligned} \tag{28}$$

It is clear that the first order Stirling interpolation formula ensures an accuracy of at least the first order Taylor series approximation. Besides, a reasonable $\hbar$ can make the remainder of the Stirling interpolation formula more closer to the high order terms of Taylor series.

On the other hand, because of the long memory property of fractional calculus, the estimated state x_k is related to all of the previous state, so the longer the memory length L, the better the filter's estimation accuracy.

Next, to analyze the numerical complexity of the proposed algorithm, the number of required floating-point operations (flops) is computed. Here, basic arithmetic operations such as addition, subtraction, multiplication, division, comparison, or square root are counted as one floating-point operation. The number of flops for vector-vector operations, matrix-vector product, and matrix-matrix product is shown in Table 1 [26].

Table 1 Computational requirements of different operations

operation	description	flops
$\mathbf{A} \pm \mathbf{B}$	$\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$	nm
$\mathbf{AB}$	$\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{B} \in \mathbb{R}^{m\times l}$	$2nml - nl$
$\mathbf{A}^{-1}$	$\mathbf{A} \in \mathbb{R}^{n\times n}$	n^3
chol($\mathbf{A}$)	Cholesky factorization, $\mathbf{A} \in \mathbb{R}^{n\times n}$	$\frac{1}{3}n^3$
	$\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{x} \in \mathbb{R}^m$	$(2m-1)n$
$\mathbf{Ax}$	$m = n$, $\mathbf{A} \in \mathbb{R}^{n\times m}$ diagonal, $\mathbf{x} \in \mathbb{R}^n$	n
	$m = n$, $\mathbf{A} \in \mathbb{R}^{n\times m}$ lower triangular, $\mathbf{x} \in \mathbb{R}^n$	$n(n+1)$

However, two nonlinear functions $\mathbf{f}(\mathbf{x}_k)$ and $\mathbf{h}(\mathbf{x}_k)$ affect the time complexity significantly, so it is hard to evaluate the exactly computational complexity. Therefore, we assume that the required flops of two functions $\mathbf{f}$ and $\mathbf{h}$ associated with the n-dimensional vector are $F(n)$ and $H(n, m)$, respectively. The specific flops of each step are shown in Table 2. Totally, the costs of the FCDKF are given by

$$\begin{aligned} T_{FCDKF} = {} & \tfrac{14}{3}n^3 + (19 + 3L)n^2 + (2L + 1)n + (4n + 1)F(n) \\ & + nm + 8nm^2 + 2n^2m + 2m + m^3 + (6n + 1)H(n, m), \end{aligned} \tag{29}$$

so $\max\{O(n^3), O(m^3), O(nF(n)), O(nH(n, m))\}$ is the numerical complexity of the proposed algorithm.

3.2. Adaptive fractional central difference Kalman filter

Assuming that $\boldsymbol{\alpha}$ represents the estimated noise parameters $\mathbf{q}$, $\mathbf{Q}$, $\mathbf{r}$, $\mathbf{R}$. In order to estimate the system state $\mathbf{x}_k$ and parameter $\boldsymbol{\alpha}$ simultaneously, based on the MAP principle [27], the AFCDKF is presented.

Table 2 Time requirements of each step

step	flops	step	flops
$\hat{\mathbf{x}}_{k\|k-1}$	$(2L+1)n + F(n)$	$\mathbf{S}_{k-1}$	$\frac{1}{3}n^3$
$\mathbf{P}_{k\|k-1}$	$4n^3 + (12+3L)n^2 + 4nF(n)$	$\mathbf{S}_{k\|k-1}$	$\frac{1}{3}n^3$
$\mathbf{P}_{\widetilde{z}}$	$4n^2 + 2nm + 2nm^2 + 4nH(n,m)$	$\mathbf{z}_{k\|k-1}$	$H(n,m) + m$
$\mathbf{P}_{\widetilde{x}\widetilde{z}}$	$2n^2 + 2n^2m + 2nH(n,m)$	$\hat{\mathbf{x}}_k$	$2nm + m$
$\mathbf{P}_k$	$n^2 + 4nm^2 - 2nm$	$\mathbf{K}_k$	$m^3 + 2nm^2 - nm$

3.2.1. Noise statistics estimator

Using the Bayesian theorem, the posterior probability density is given by

$$p[\mathbf{X}_k, \boldsymbol{\alpha} \mid \mathbf{Z}_k] = \frac{p[\mathbf{X}_k, \boldsymbol{\alpha}, \mathbf{Z}_k]}{p[\mathbf{Z}_k]}, \tag{30}$$

where $\mathbf{X}_k = [\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_k]$ and $\mathbf{Z}_k = [\mathbf{z}_0, \mathbf{z}_1, \cdots, \mathbf{z}_k]$.

$p[\mathbf{Z}_k]$ is unrelated to the parameters and system state. Thus the problem can be transformed into optimizing the following objective function

$$J = p[\mathbf{X}_k, \boldsymbol{\alpha}, \mathbf{Z}_k] = p[\mathbf{Z}_k \mid \mathbf{X}_k, \boldsymbol{\alpha}]\; p[\mathbf{X}_k \mid \boldsymbol{\alpha}]\; p[\boldsymbol{\alpha}], \tag{31}$$

where $p[\boldsymbol{\alpha}]$ can be considered as a constant.

Exploiting Assumptions 2.3 and 2.6, $p[\mathbf{X}_k \mid \boldsymbol{\alpha}]$ is formulated as

$$
\begin{aligned}
p[\mathbf{X}_k \mid \boldsymbol{\alpha}] &= p[\mathbf{x}_0 \mid \boldsymbol{\alpha}]\ p[\mathbf{x}_1 \mid \mathbf{x}_0, \boldsymbol{\alpha}]\ p[\mathbf{x}_2 \mid \mathbf{X}_1, \boldsymbol{\alpha}]\ \cdots\ p[\mathbf{x}_k \mid \mathbf{X}_{k-1}, \boldsymbol{\alpha}] \\
&= \tfrac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{P}_0|^{\frac{1}{2}}} \exp[-\tfrac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^{\mathrm{T}} \mathbf{P}_0^{-1} (\mathbf{x}_0 - \hat{\mathbf{x}}_0)] \prod_{j=1}^{k} \Big\{ \tfrac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{Q}|^{\frac{1}{2}}} \\
&\quad \times \exp\{-\tfrac{1}{2}[\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j} (-1)^i \boldsymbol{\gamma}_i \mathbf{x}_{j-i} - \mathbf{q}]^{\mathrm{T}} \mathbf{Q}^{-1} \\
&\quad \times [\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j} (-1)^i \boldsymbol{\gamma}_i \mathbf{x}_{j-i} - \mathbf{q}]\}\Big\} \\
&= \tfrac{1}{(2\pi)^{\frac{n(k+1)}{2}} |\mathbf{P}_0|^{\frac{1}{2}} |\mathbf{Q}|^{\frac{k}{2}}} \exp\Big\{ -\tfrac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^{\mathrm{T}} \mathbf{P}_0^{-1} (\mathbf{x}_0 - \hat{\mathbf{x}}_0) \\
&\quad - \tfrac{1}{2} \textstyle\sum_{j=1}^{k} \{[\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \sum_{i=1}^{j} (-1)^i \boldsymbol{\gamma}_i \mathbf{x}_{j-i} - \mathbf{q}]^{\mathrm{T}} \mathbf{Q}^{-1} \\
&\quad \times [\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j} (-1)^i \boldsymbol{\gamma}_i \mathbf{x}_{j-i} - \mathbf{q}]\}\Big\},
\end{aligned} \tag{32}
$$

where $|\cdot|$ means to the determinant, denoted as $|A| = \det(A)$.

Similarly, $p[\mathbf{Z}_k \mid \mathbf{X}_k, \boldsymbol{\alpha}]$ can be updated as

$$
\begin{aligned}
& p[\mathbf{Z}_k \mid \mathbf{X}_k, \boldsymbol{\alpha}] \\
&= \prod_{j=1}^{k} p[\mathbf{z}_j \mid \mathbf{X}_k, \boldsymbol{\alpha}] \\
&= \prod_{j=1}^{k} \tfrac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\Big\{ -\tfrac{1}{2}[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]^{\mathrm{T}} \mathbf{R}^{-1} [\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]\Big\} \\
&= \tfrac{1}{(2\pi)^{\frac{mk}{2}} |\mathbf{R}|^{\frac{k}{2}}} \exp\Big\{ -\tfrac{1}{2} \textstyle\sum_{j=1}^{k} \{[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]^{\mathrm{T}} \mathbf{R}^{-1} [\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]\}\Big\}.
\end{aligned} \tag{33}
$$

Employing (32) and (33) into (31), the objective function J can be reformulated as

$$
\begin{aligned}
J &= p[\mathbf{Z}_k \mid \mathbf{X}_k, \boldsymbol{\alpha}]\ p[\mathbf{X}_k \mid \boldsymbol{\alpha}]\ p[\boldsymbol{\alpha}] \\
&= \tfrac{1}{(2\pi)^{\frac{(n+m)k+n}{2}} |\mathbf{P}_0|^{\frac{1}{2}} |\mathbf{Q}|^{\frac{k}{2}} |\mathbf{R}|^{\frac{k}{2}}} \exp[-\tfrac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^{\mathrm{T}} \mathbf{P}_0^{-1} (\mathbf{x}_0 - \hat{\mathbf{x}}_0)] \\
&\quad \times \exp\Big\{ -\tfrac{1}{2} \textstyle\sum_{j=1}^{k} \{[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]^{\mathrm{T}} \mathbf{R}^{-1} [\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}] \\
&\quad + [\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j} (-1)^i \boldsymbol{\gamma}_i \mathbf{x}_{j-i} - \mathbf{q}]^{\mathrm{T}} \mathbf{Q}^{-1} \\
&\quad \times [\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j} (-1)^i \boldsymbol{\gamma}_i \mathbf{x}_{j-i} - \mathbf{q}]\}\Big\}\ p[\boldsymbol{\alpha}].
\end{aligned} \tag{34}
$$

Maximizing the objective function J is equivalent to maximizing $\ln J$, and then the objective function can be reformulated as

$$\begin{aligned}\ln J = -\tfrac{1}{2}\textstyle\sum_{j=1}^{k}\Big\{&[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}] - \tfrac{k}{2}\ln|\mathbf{Q}| \\ &-\tfrac{k}{2}\ln|\mathbf{R}| + [\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\mathbf{x}_{j-i} - \mathbf{q}]^{\mathrm{T}}\mathbf{Q}^{-1} \\ &\times[\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\mathbf{x}_{j-i} - \mathbf{q}]\Big\} + \mathrm{C},\end{aligned} \tag{35}$$

where $\mathrm{C} = -\frac{1}{2}\ln|\mathbf{P}_0| - \frac{(n+m)k+n}{2}\ln\frac{1}{2\pi} + \ln p[\boldsymbol{\alpha}] - \frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^{\mathrm{T}}\mathbf{P}_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)$ is a constant.

As mentioned before, the estimated parameter $\boldsymbol{\alpha}$ includes $\mathbf{q}$, $\mathbf{Q}$, $\mathbf{r}$, $\mathbf{R}$. Let

$$\frac{\partial \ln J}{\partial \boldsymbol{\alpha}} = 0. \tag{36}$$

The following equations can be obtained.

$$\tfrac{\partial \ln J}{\partial \mathbf{q}} = \tfrac{1}{2}\textstyle\sum_{j=1}^{k}\mathbf{Q}^{-1}[\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\mathbf{x}_{j-i} - \mathbf{q}] = 0, \tag{37}$$

$$\begin{aligned}\tfrac{\partial \ln J}{\partial \mathbf{Q}} &= \tfrac{1}{2}(\mathbf{Q}^{\mathrm{T}})^{-1}\textstyle\sum_{j=1}^{k}[\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\mathbf{x}_{j-i} - \mathbf{q}] \\ &\quad\times[\mathbf{x}_j - \mathbf{f}(\mathbf{x}_{j-1}) + \textstyle\sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\mathbf{x}_{j-i} - \mathbf{q}]^{\mathrm{T}}(\mathbf{Q}^{\mathrm{T}})^{-1} - \tfrac{k}{2}(\mathbf{Q}^{\mathrm{T}})^{-1} \\ &= 0,\end{aligned} \tag{38}$$

$$\tfrac{\partial \ln J}{\partial \mathbf{r}} = \tfrac{1}{2}\mathbf{R}^{-1}\textstyle\sum_{j=1}^{k}[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}] = 0, \tag{39}$$

$$\begin{aligned}\tfrac{\partial \ln J}{\partial \mathbf{R}} &= \tfrac{1}{2}(\mathbf{R}^{\mathrm{T}})^{-1}\textstyle\sum_{j=1}^{k}[\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}][\mathbf{z}_j - \mathbf{h}(\mathbf{x}_j) - \mathbf{r}]^{\mathrm{T}}(\mathbf{R}^{\mathrm{T}})^{-1} - \tfrac{k}{2}(\mathbf{R}^{\mathrm{T}})^{-1} \\ &= 0.\end{aligned} \tag{40}$$

In the previous formulas, the real state and parameters cannot be obtained, so the estimated $\hat{\mathbf{x}}_j$, $\hat{\mathbf{x}}_{j|j-1}$ are employed to replace the real value $\mathbf{x}_j$. Then, the estimated noise parameters are given by

$$\hat{\mathbf{q}}_k = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}[\hat{\mathbf{x}}_j - \mathbf{f}(\hat{\mathbf{x}}_{j-1}) + \sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\hat{\mathbf{x}}_{j-i}], \tag{41}$$

$$\begin{aligned}\hat{\mathbf{Q}}_k = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}&[\hat{\mathbf{x}}_j - \mathbf{f}(\hat{\mathbf{x}}_{j-1}) + \textstyle\sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\hat{\mathbf{x}}_{j-i} - \mathbf{q}]\\ &\times[\hat{\mathbf{x}}_j - \mathbf{f}(\hat{\mathbf{x}}_{j-1}) + \textstyle\sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\hat{\mathbf{x}}_{j-i} - \mathbf{q}]^{\mathrm{T}},\end{aligned} \tag{42}$$

$$\hat{\mathbf{r}}_k = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}[\mathbf{z}_j - \mathbf{h}(\hat{\mathbf{x}}_{j|j-1})], \tag{43}$$

$$\hat{\mathbf{R}}_k = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}[\mathbf{z}_j - \mathbf{h}(\hat{\mathbf{x}}_{j|j-1}) - \mathbf{r}][\mathbf{z}_j - \mathbf{h}(\hat{\mathbf{x}}_{j|j-1}) - \mathbf{r}]^{\mathrm{T}}. \tag{44}$$

3.2.2. Unbiased analysis

As we can see, the noise parameters can be obtained by solving (41)–(44). For integer order nonlinear systems with accurate posterior information, it has been proved that the output error $\boldsymbol{\varepsilon}_j = \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}$ subjects to zero-mean Gaussian white noise sequence [27, 25]. Similar conclusions can be generalized to systems described by Definition 2.1. Then

$$\begin{aligned}\mathrm{E}\{\hat{\mathbf{q}}_k\} &= \tfrac{1}{k}\textstyle\sum_{j=1}^{k}\mathrm{E}\{\hat{\mathbf{x}}_j - \mathbf{f}(\hat{\mathbf{x}}_{j-1}) + \textstyle\sum_{i=1}^{j}(-1)^i\boldsymbol{\gamma}_i\hat{\mathbf{x}}_{j-i}\}\\ &= \tfrac{1}{k}\textstyle\sum_{j=1}^{k}\mathbf{K}_j\mathrm{E}\{\boldsymbol{\varepsilon}_j\} + \mathbf{q}\\ &= \mathbf{q},\end{aligned} \tag{45}$$

$$\mathrm{E}\{\hat{\mathbf{r}}_k\} = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}\mathrm{E}\{\mathbf{z}_j - \mathbf{h}(\hat{\mathbf{x}}_{j|j-1})\} = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}\mathrm{E}\{\boldsymbol{\varepsilon}_j\} + \mathbf{r} = \mathbf{r}. \tag{46}$$

Therefore, $\hat{\mathbf{q}}_k$ and $\hat{\mathbf{r}}_k$ are unbiased. Next, the unbiased analysis of noise covariance estimation $\hat{\mathbf{Q}}_k$ and $\hat{\mathbf{R}}_k$ is discussed. Employing (11) and (24) into (42) yields

$$\mathrm{E}\{\hat{\mathbf{Q}}_k\} = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}\mathbf{K}_j\mathrm{E}\{\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}\}\mathbf{K}_j^{\mathrm{T}} = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}\mathbf{K}_j\mathbf{P}_{\tilde{\mathbf{z}}_j}\mathbf{K}_j^{\mathrm{T}}, \tag{47}$$

and substituting (20), (22) and (26) into (47) results in

$$\mathrm{E}\{\hat{\mathbf{Q}}_k\} = \tfrac{1}{k}\textstyle\sum_{j=1}^{k}(\mathbf{P}_{j|j-1} - \mathbf{P}_j) \neq \mathbf{Q}. \tag{48}$$

The measurement noise covariance estimation $\hat{\mathbf{R}}_k$ can be formulated as

$$\begin{aligned} \mathrm{E}\{\hat{\mathbf{R}}_k\} &= \frac{1}{k}\sum_{j=1}^{k}\mathrm{E}\{[\mathbf{z}_j-\mathbf{h}(\hat{\mathbf{x}}_{j|j-1})-\mathbf{r}][\mathbf{z}_j-\mathbf{h}(\hat{\mathbf{x}}_{j|j-1})-\mathbf{r}]^{\mathrm{T}}\} \\ &= \frac{1}{k}\sum_{j=1}^{k}\mathrm{E}\{\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}\}. \end{aligned} \tag{49}$$

Utilizing (22), (49) can be reformulated as

$$\begin{aligned} \mathrm{E}\{\hat{\mathbf{R}}_k\} = \frac{1}{k}\sum_{j=1}^{k}\Big\{ & \frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}+\hbar\mathbf{s}^i_{j|j-1})-\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}-\hbar\mathbf{s}^i_{j|j-1})] \\ & \times[\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}+\hbar\mathbf{s}^i_{j|j-1})-\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}-\hbar\mathbf{s}^i_{j|j-1})]^{\mathrm{T}}+\mathbf{R}\Big\} \\ \neq \mathbf{R}. & \end{aligned} \tag{50}$$

Then we can obtain the unbiased noise covariance estimation directly,

$$\begin{aligned} \hat{\mathbf{Q}}_k = \frac{1}{k}\sum_{j=1}^{k}\Big\{ & K_j\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}K_j^{\mathrm{T}}-\frac{1}{\hbar}\mathbf{G}^{j-1}_{\tilde{\mathbf{f}}}\mathbf{S}^{\mathrm{T}}_{j-1}\boldsymbol{\gamma}_1^{\mathrm{T}}-\frac{1}{\hbar}\boldsymbol{\gamma}_1\mathbf{S}_{j-1}(\mathbf{G}^{j-1}_{\tilde{\mathbf{f}}})^{\mathrm{T}} \\ & -\frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{f}(\hat{\mathbf{x}}_{j-1}+\hbar\mathbf{s}^i_{j-1})-\mathbf{f}(\hat{\mathbf{x}}_{j-1}-\hbar\mathbf{s}^i_{j-1})] \\ & \times[\mathbf{f}(\hat{\mathbf{x}}_{j-1}+\hbar\mathbf{s}^i_{j-1})-\mathbf{f}(\hat{\mathbf{x}}_{j-1}-\hbar\mathbf{s}^i_{j-1})]^{\mathrm{T}} \\ & -\sum_{m=1}^{j}\boldsymbol{\gamma}_m\mathbf{P}_{j-m}\boldsymbol{\gamma}_m^{\mathrm{T}}+\mathbf{P}_j\Big\}, \end{aligned} \tag{51}$$

$$\begin{aligned} \hat{\mathbf{R}}_k = \frac{1}{k}\sum_{j=1}^{k}\Big\{ & -\frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}+\hbar\mathbf{s}^i_{j|j-1})-\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}-\hbar\mathbf{s}^i_{j|j-1})] \\ & \times[\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}+\hbar\mathbf{s}^i_{j|j-1})-\mathbf{h}(\hat{\mathbf{x}}_{j|j-1}-\hbar\mathbf{s}^i_{j|j-1})]^{\mathrm{T}}+\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}}\Big\}. \end{aligned} \tag{52}$$

To further reduce the computation complexity, the recursive formulas are developed.

$$\hat{\mathbf{q}}_k = \frac{1}{k}[(k-1)\hat{\mathbf{q}}_{k-1}+\hat{\mathbf{x}}_k-\mathbf{f}(\hat{\mathbf{x}}_{k-1})+\sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j}], \tag{53}$$

$$\begin{aligned} \hat{\mathbf{Q}}_k = \frac{1}{k}\Big\{ & (k-1)\hat{\mathbf{Q}}_{k-1}-\frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{f}(\hat{\mathbf{x}}_{k-1}+\hbar\mathbf{s}^i_{k-1})-\mathbf{f}(\hat{\mathbf{x}}_{k-1}-\hbar\mathbf{s}^i_{k-1})] \\ & \times[\mathbf{f}(\hat{\mathbf{x}}_{k-1}+\hbar\mathbf{s}^i_{k-1})-\mathbf{f}(\hat{\mathbf{x}}_{k-1}-\hbar\mathbf{s}^i_{k-1})]^{\mathrm{T}}-\sum_{m=1}^{k}\boldsymbol{\gamma}_m\mathbf{P}_{k-m}\boldsymbol{\gamma}_m^{\mathrm{T}} \\ & +K_k\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}}K_k^{\mathrm{T}}-\frac{1}{\hbar}\mathbf{G}^{k-1}_{\tilde{\mathbf{f}}}\mathbf{S}^{\mathrm{T}}_{k-1}\boldsymbol{\gamma}_1^{\mathrm{T}}-\frac{1}{\hbar}\boldsymbol{\gamma}_1\mathbf{S}_{k-1}(\mathbf{G}^{k-1}_{\tilde{\mathbf{f}}})^{\mathrm{T}}+\mathbf{P}_k\Big\}, \end{aligned} \tag{54}$$

$$\hat{\mathbf{r}}_k = \frac{1}{k}[(k-1)\hat{\mathbf{r}}_{k-1} + \mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})], \tag{55}$$

$$\begin{aligned}\hat{\mathbf{R}}_k = \frac{1}{k}\Big\{&(k-1)\hat{\mathbf{R}}_{k-1} + \boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}} \\ &-\frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}^i_{k|k-1}) - \mathbf{h}(\hat{\mathbf{x}}_{k-1} - \hbar\mathbf{s}^i_{k|k-1}) \\ &\times[\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}^i_{k|k-1}) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar\mathbf{s}^i_{k|k-1})]^{\mathrm{T}}\Big\}.\end{aligned} \tag{56}$$

Finally, the AFCDKF with unknown prior knowledge is derived. The proposed AFCDKF can evaluate state and noise parameters simultaneously. The recursively pseudocode is shown in Algorithm 2.

Algorithm 2 Adaptive fractional central difference Kalman filter

Initialization:

Set the system initial values: $\mathbf{x}_0, \mathbf{P}_0$

Set the noise stochastic initial values: $\mathbf{q}_0, \mathbf{r}_0, \mathbf{Q}_0, \mathbf{R}_0$

Set the short memory principle length: L

Set the interval length: $\hbar$

On-line updating:

for $i = 1 \to k-1$ **do**

calculate Algorithm *1* on-line updating part

Noise statistics estimator:

$$\hat{\mathbf{q}}_k = \frac{1}{k}[(k-1)\hat{\mathbf{q}}_{k-1} + \hat{\mathbf{x}}_k - \mathbf{f}(\hat{\mathbf{x}}_{k-1}) + \sum_{j=1}^{k}(-1)^j\boldsymbol{\gamma}_j\hat{\mathbf{x}}_{k-j}]$$

▷ estimated system noise mean

$$\hat{\mathbf{r}}_k = \frac{1}{k}[(k-1)\hat{\mathbf{r}}_{k-1} + \mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})]$$

▷ estimated measurement noise mean

$$\begin{aligned}\hat{\mathbf{Q}}_k = \frac{1}{k}\Big\{&(k-1)\hat{\mathbf{Q}}_{k-1} + \mathbf{P}_k + K_k\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}}K_k^{\mathrm{T}} - \frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{f}(\hat{\mathbf{x}}_{k-1} + \hbar\mathbf{s}^i_{k-1}) \\ &-\mathbf{f}(\hat{\mathbf{x}}_{k-1} - \hbar\mathbf{s}^i_{k-1})][\mathbf{f}(\hat{\mathbf{x}}_{k-1} + \hbar\mathbf{s}^i_{k-1}) - \mathbf{f}(\hat{\mathbf{x}}_{k-1} - \hbar\mathbf{s}^i_{k-1})]^{\mathrm{T}} \\ &-\frac{1}{\hbar}\mathbf{G}^{k-1}_{\tilde{\mathbf{f}}}\mathbf{S}^{\mathrm{T}}_{k-1}\boldsymbol{\gamma}_1^{\mathrm{T}} - \frac{1}{\hbar}\boldsymbol{\gamma}_1\mathbf{S}_{k-1}(\mathbf{G}^{k-1}_{\tilde{\mathbf{f}}})^{\mathrm{T}} - \sum_{m=1}^{k}\boldsymbol{\gamma}_m\mathbf{P}_{k-m}\boldsymbol{\gamma}_m^{\mathrm{T}}\Big\}\end{aligned}$$

▷ estimated system noise covariance

$$\hat{\mathbf{R}}_k = \frac{1}{k}\Big\{(k-1)\hat{\mathbf{R}}_{k-1} + \boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^{\mathrm{T}} - \frac{1}{4\hbar^2}\sum_{i=1}^{n}[\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}_{k|k-1}^i) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar\mathbf{s}_{k|k-1}^i)][\mathbf{h}(\hat{\mathbf{x}}_{k|k-1} + \hbar\mathbf{s}_{k|k-1}^i) - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} - \hbar\mathbf{s}_{k|k-1}^i)]^{\mathrm{T}}\Big\}$$

▷ estimated measurement noise covariance

8: **end for**

Remark 3.3 *The numerical complexity of the AFCDKF is the same as the proposed FCDKF. Without sacrificing time complexity, the merits of the p-resented AFCDKF are summarized below*

- *Using the Stirling interpolation formula, the proposed FCDKF can estimate the system state as no derivatives are needed. When the order* $\alpha = 1$, *the FCDKF can be reduced to the CDKF.*
- *Based on the MAP principle, the AFCDKF algorithm can estimate parameters and state concurrently. Besides, the MAP principle based parameter estimation algorithm is unbiased.*

Remark 3.4 *The MAP principle based AFCDKF algorithm can estimate parameters unbiasedly. In the future, we will further improve the proposed adaptive algorithm to evaluate the noise covariance matrixes* Q *and* R *simultaneously.*

4. Illustrative Examples

To demonstrate the performance of the FCDKF and AFCDKF, several fractional discrete nonlinear dynamic plants are considered, including a s-calar system and a multidimensional system. All algorithms are coded with MATLAB R2017a. The simulations are carried out on a computer with Intel

Core i3-2350M CPU @2.30GHz and RAM with 8.00 GB. Besides, the running time of program is computed by *tic* and *toc* in MATLAB to measure the algorithm performance.

4.1. Scalar system

Consider the following scalar system, whose state space model can be represented by

$$\begin{cases} \nabla^{0.7} x_k = 3\sin(2x_{k-1}) - x_{k-1} + \omega_k, \\ y_k = x_k + \nu_k. \end{cases} \tag{57}$$

System state x_k and measurement value y_k are polluted by system noise ω_k and measurement noise ν_k respectively. ω_k and ν_k are uncorrelated Gaussian noise.

4.1.1. Fractional central difference Kalman filter

In this subsection, the proposed FCDKF is used to implement an on-line real-time state estimation. Considering the plant mentioned before, noise distributions are respectively selected as $\boldsymbol{\omega}_k \sim \mathcal{N}(1, 0.81)$ and $\boldsymbol{\nu}_k \sim \mathcal{N}(1, 0.25)$. The algorithm parameters are set as: the initial state $x_0 = 0$, the initial covariance matrix $P_0 = 100$, the short memory principle length $L = 10$, the interval length $\hbar = \sqrt{3}$.

To evaluate the accuracy of state, the square error (SE) and the root mean square error (RMSE) are selected as performance indexes, which are described as

$$\begin{cases} \text{SE} \triangleq (x_k - \hat{x}_k)^2, \\ \text{RMSE} \triangleq \sqrt{\frac{1}{k}\sum_{j=1}^{k}(x_j - \hat{x}_j)^2}. \end{cases} \tag{58}$$

The estimated state of the fractional system is shown in Fig. 1. It can be observed that the proposed FCDKF exhibits good performance in the presence of noise.

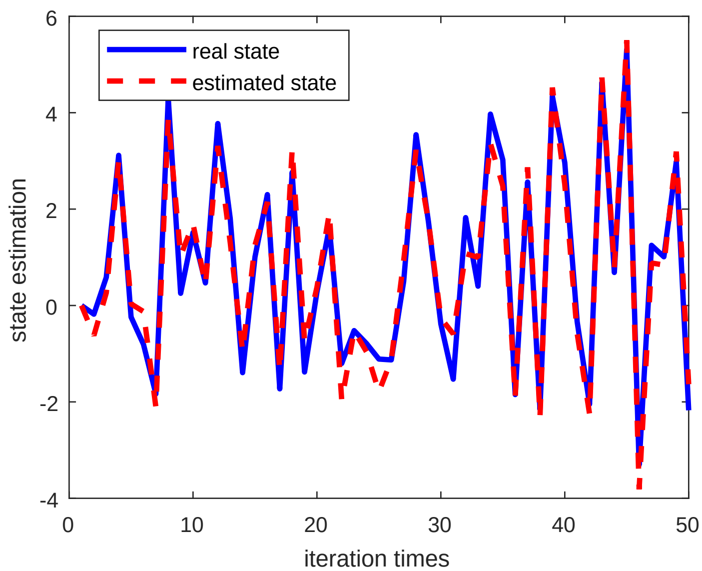

Fig. 1 State estimation of the FCDKF

Then, the effectiveness of the proposed FCDKF, FPF and FEKF presented in [21] is compared. Fig. 2 shows SE of the FCDKF, FEKF and FPF. To enhance the persuasion, 50 Monte Carlo experiments are conducted and the corresponding results are presented in Table 3, where N indicates the number of particles and estimation error $e = x - \hat{x}$. From the simulation results (Fig. 2 and Table 3), we can obtain that the proposed FCDKF performs better estimation performance. Moreover, compared with the FPF, the FCDKF performs better in terms of estimation accuracy and real-time.

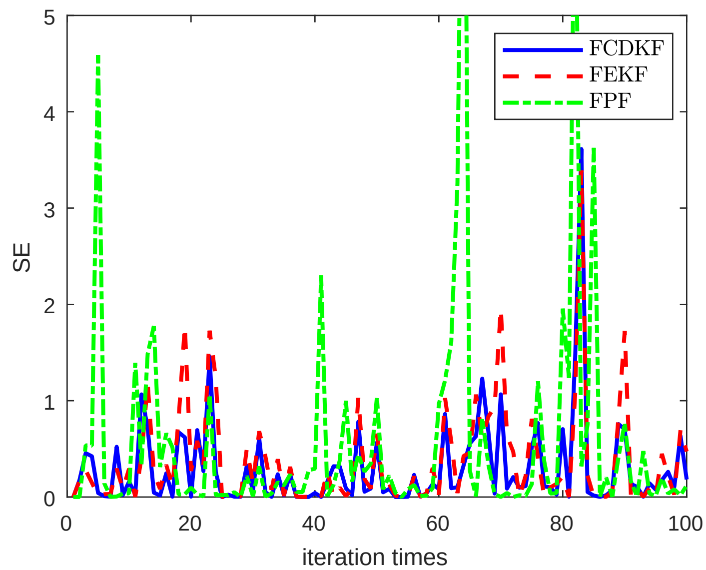

Fig. 2 Comparison between the FCDKF, FEKF and FPF

4.1.2. Adaptive fractional central difference Kalman filter

Next, the validity of the AFCDKF discussed in Section 3 is investigated. Considering the aforementioned plant, the real noise distributions are chosen as $\boldsymbol{\omega}_k \sim \mathcal{N}(6, 10)$ and $\boldsymbol{\nu}_k \sim \mathcal{N}(5, 0.25)$, the algorithm parameters are set as: the initial state $x_0 = 0$, the initial covariance matrix $P_0 = 100$ and the short memory principle length $L = 10$. The interval length is altered to $\hbar = \sqrt{1.3}$.

First the convergence of the parameter estimation is verified. The results are shown in Fig. 3. The two parameter estimation curves converge to the real values with iteration times. The experimental results show that the unbiased estimation by using the proposed method is indeed obtained.

To clarify the stochastic property of the proposed algorithm, we do 1000 Monte Carlo experiments, and estimated parameters are shown in Fig. 4.

Table 3 Performance analysis of algorithms

	N	running time (s)	$\|e\|_1$	$\|e\|_2$
FCDKF	-	0.2162	39.6251	4.9930
FEKF	-	0.2080	40.1034	5.0530
FPF	50	0.2268	40.7246	5.4080
	100	0.2635	39.0319	5.0036
	200	0.3408	38.5439	4.8978

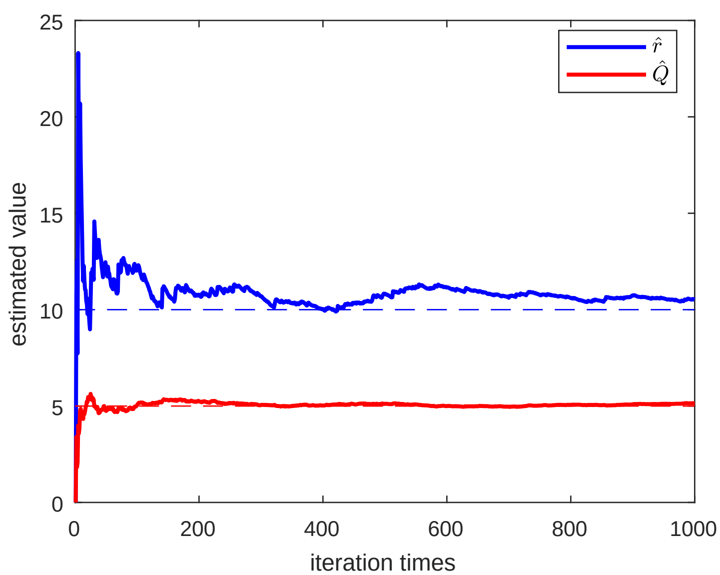

Fig. 3 Convergence of estimated parameters

The red dot represents real parameters, the histogram below the scatter plot represents the kernel density of the measurement noise mean r, and the left indicates the kernel density of the system noise covariance Q. As we can see, the estimated mean $\hat{r}$ concentrates on the side of the true value, and the estimated covariance $\hat{Q}$ deviates from the true value slightly. Therefore the effectiveness of the AFCDKF is confirmed.

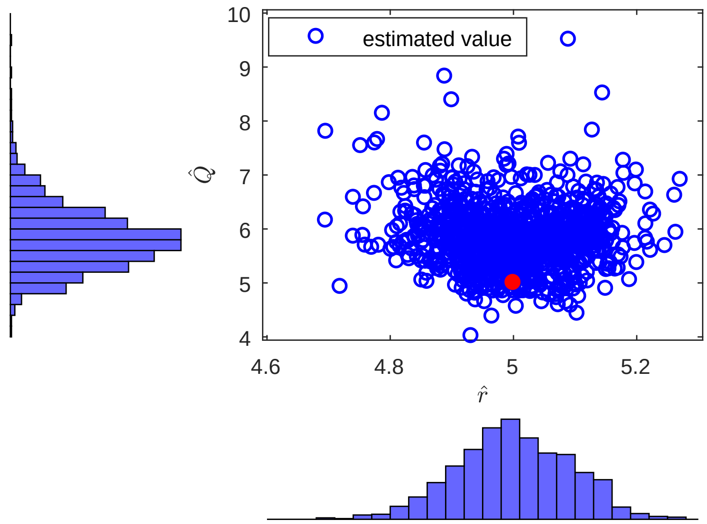

Fig. 4 Estimated parameters in 1000 trials

Next, the state estimation accuracy of the AFCDKF is investigated. Assuming that the system noise covariance Q and the measurement noise mean r are unknown, the AFCDKF is utilized to evaluate the state and parameters simultaneously. For comparison, the proposed FCDKF is also employed to evaluate the same system state. For the FCDKF, due to the real parameters r and Q are unknown, the system noise mean and the measurement noise covariance employed in the FCDKF are set as 4 and 8, respectively. The SE is shown in Fig. 5. When the prior information is unknown, the figure exhibits that the AFCDKF outperforms the FCDKF.

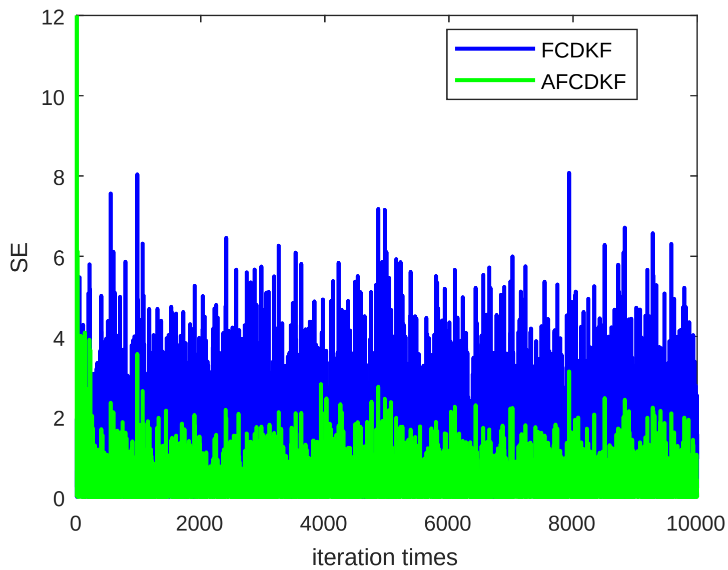

Fig. 5 Estimation errors of the FCDKF and the AFCDKF

4.2. Multidimensional system

In order to further verify the effectiveness of two algorithms, a multidimensional system is considered

$$\begin{cases} \nabla^{\boldsymbol{\alpha}} \boldsymbol{x}_k = \begin{bmatrix} \cos(x_{2,k-1}) \\ -0.1x_{2,k-1} + \mathrm{e}^{-0.05x_{3,k-1}} + u_{k-1} \\ -x_{3,k-1} - 0.5|x_{1,k-1}| \end{bmatrix} + \boldsymbol{\omega}_k, \\ y_k = 0.1x_{1,k} + 0.2x_{2,k} + \nu_k, \end{cases} \tag{59}$$

where $\boldsymbol{x}_k = [x_{1,k}, x_{2,k}, x_{3,k}]^{\mathrm{T}}$.

4.2.1. Fractional central difference Kalman filter

The algorithm parameters are set as: fractional orders $\boldsymbol{\alpha} = [0.7, 1.2, 0.5]^{\mathrm{T}}$, $\mathbf{q} = [0, 0, 0]^{\mathrm{T}}$, $r = 0$, $\mathbf{Q} = \mathrm{diag}[0.3, 0.3, 0.001]$, $R = 0.3$, $\hbar = \sqrt{3}$, the initial real state $\mathbf{x}_0 = [0, 0, 0.2]^{\mathrm{T}}$, the initial covariance matrix $\mathbf{P}_0 = \mathrm{diag}[100, 100, 100]$, the initial estimation state $\hat{\mathbf{x}}_0 = [0.1, 0.1, 0.1]^{\mathrm{T}}$. Besides, the input signal is the Gaussian random white noise with the distribution $\mathcal{N}(0, 1)$.

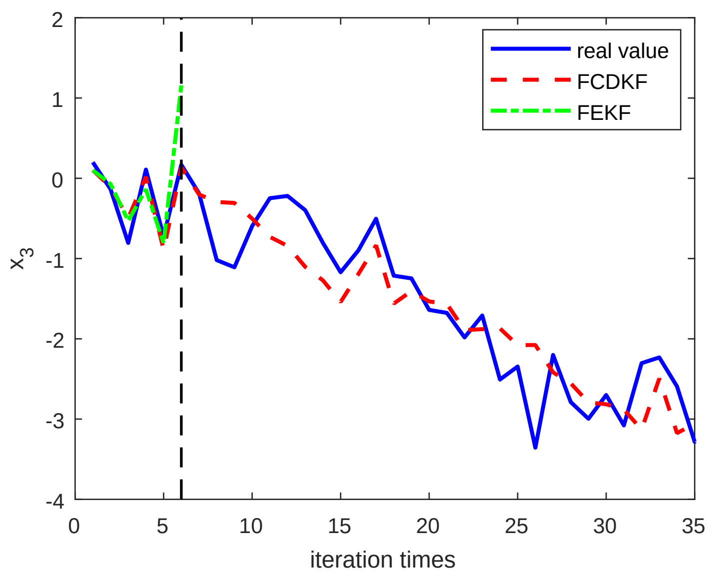

Fig. 6 Comparison between the FCDKF and FEKF

According to (59), it is clear that the system function is continuous and non-differentiable at $x_{1,k} = 0$. From the simulation results (Fig. 6 and Table 4), when the estimated state $\hat{x}_{1,6} = 0$, the Jacobian matrix of the system function does not exist and several unexpected values (e.g. NaN) appear, so the FEKF is out of effect. For the proposed FCDKF, it can still estimate the system state effectively because it is a derivative-free filtering algorithm. Therefore the effectiveness of the FCDKF is confirmed.

Remark 4.1 *The numerical accuracy of MATLAB is so high that situations where the state equals to 0 will hardly appear. However, most of algorithms*

Table 4 The partial state estimation results of a multidimensional system

		iteration time k			
state	$\mathbf{x}$	5	6	7	8
real value	x_1	0.2648	–0.0285	–2.1050	–3.0517
	x_2	3.5304	3.1571	3.0850	4.6176
	x_3	–0.7253	0.1687	–0.1870	–1.0194
FCDKF	$\hat{x}_1$	0.5073	0	0.6444	0.7759
	$\hat{x}_2$	2.8416	5.2236	5.2380	5.1037
	$\hat{x}_3$	–0.8909	0.1342	–0.2010	–0.2951
FEKF	$\hat{x}_1$	0.2086	0	NaN	NaN
	$\hat{x}_2$	2.9267	5.4280	NaN	NaN
	$\hat{x}_3$	–0.8086	1.1573	NaN	NaN

are carried out on microcontrollers with limited precision, so in this experiment, when a state value is less than 0.01, the state value is set as 0.

4.2.2. Adaptive fractional central difference Kalman filter

Next, the AFCDKF for a multidimensional system is also considered. The algorithm parameters remain the same. Assuming the measurement noise expectation r is unknown, the initial estimated value of measurement noise expectation $\hat{r}_0$ is set as 0.5. As shown in Fig. 7, the estimated parameter $\hat{r}$ converges to the real value with iteration times. Besides, for comparison, the FCDKF is also utilized to estimate the system state. The RMSE is shown in Fig. 8. When the noise prior information is unknown, the AFCDKF performs better than FCDKF.

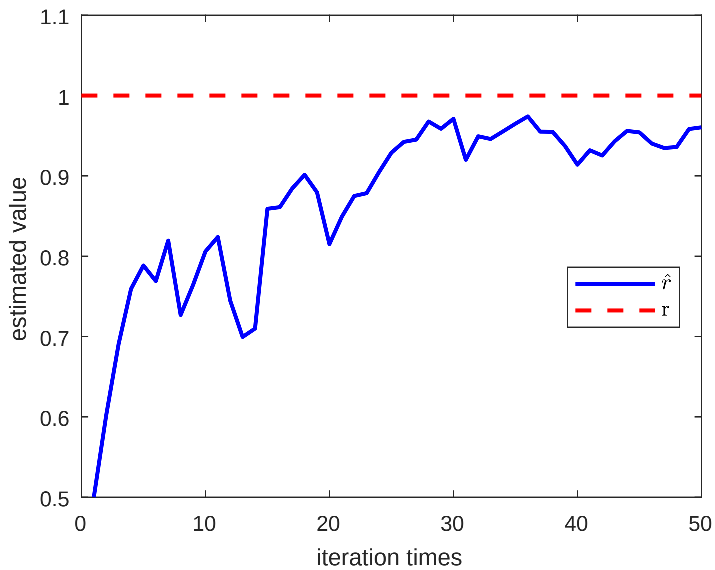

Fig. 7 Parameter estimation for the MIMO system

5. Conclusions

In this paper, the state estimation problem is investigated for nonlinear discrete fractional dynamic systems. Two filtering algorithms have been developed. The FCDKF algorithm can be implemented without derivative signal. Furthermore, a recursive AFCDKF is achieved, which can evaluate the parameters and state simultaneously. The approximate accuracy and numerical complexity of the two algorithms are analyzed. Effectiveness of the proposed algorithms is illustrated through several simulation examples where the FCDKF has superior estimation performance and the AFCDKF gives the unbiased parameters estimation. In addition, as the prior information is unknown, the AFCDKF outperforms the FCDKD.

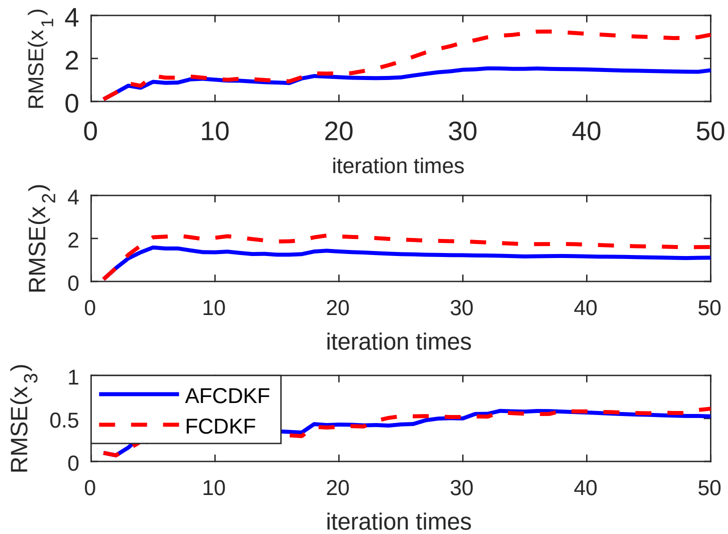

Fig. 8 Comparison between the AFCDKF and FCDKF

Acknowledgements

The authors would like to thank the Associate Editor and the anonymous reviewers for their keen and insightful comments which greatly improved the contents and presentation. The work described in this paper was fully supported by the National Natural Science Foundation of China (No. 61573332, No. 61601431), the Fundamental Research Funds for the Central Universities (No. WK2100100028), the Anhui Provincial Natural Science Foundation (No. 1708085QF141) and the General Financial Grant from the China Postdoctoral Science Foundation (No. 2016M602032).

References

[1] G. Y. Kulikov, M. V. Kulikova, Accurate continuous–discrete unscented Kalman filtering for estimation of nonlinear continuous–time stochastic models in radar tracking, *Signal Processing* 139 (2017) 25–35.

[2] M. Uricár, V. Franc, V. Hlavác, Facial landmark tracking by tree–based deformable part model based detector, in: *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW 2015)*, Santiago, 2015, pp. 10–17.

[3] Y. Demitrit, S. Verling, T. Stastny, A. Melzer, R. Siegwart, Model–based wind estimation for a hovering VTOL tailsitter UAV, in: *2017 IEEE International Conference on Robotics and Automation (ICRA 2017)*, Singapore, 2017, pp. 3945–3952.

[4] R. E. Kalman, A new approach to linear filtering and prediction problems, *Transactions of the ASME–Journal of Basic Engineering* 82 (Series D) (1960) 35–45.

[5] Y. Sunahara, An approximate method of state estimation for nonlinear dynamical systems, *Transactions of the ASME–Journal of Basic Engineering* 92 (2) (1970) 385–393.

[6] R. S. Bucy, K. D. Senne, Digital synthesis of non–linear filters, *Automatica* 7 (3) (1971) 287–298.

[7] S. Julier, J. Uhlmann, H. F. Durrant-Whyte, A new method for the nonlinear transformation of means and covariances in filters and estimators, *IEEE Transactions on Automatic Control* 45 (3) (2000) 477–482.

[8] S. K. Biswas, L. Qiao, A. G. Dempster, A novel a priori state computation strategy for the unscented Kalman filter to improve computational efficiency, *IEEE Transactions on Automatic Control* 62 (4) (2017) 1852–1864.

[9] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problems, *IEEE Transactions on Automatic Control* 45 (5) (2000) 910–927.

[10] M. NøRgaard, N. K. Poulsen, O. Ravn, New developments in state estimation for nonlinear systems, *Automatica* 36 (11) (2000) 1627–1638.

[11] V. D. M. Rudolph, Sigma–point Kalman filters for probabilistic inference in dynamic state–space models, Ph.D. thesis, Oregon Health & Science University, Oregon (2004).

[12] R. Mehra, Approaches to adaptive filtering, *IEEE Transactions on Automatic Control* 17 (5) (1972) 693–698.

[13] A. P. Sage, G. W. Husa, Adaptive filtering with unknown prior statistics, in: *Joint Automatic Control Conference*, no. 7, Colorado, 1969, pp. 760–769.

[14] Y. L. Huang, Y. G. Zhang, Z. M. Wu, N. Li, J. Chambers, A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices, *IEEE Transactions on Automatic Control* 63 (2) (2018) 594–601.

[15] Y. H. Wei, Y. Q. Chen, S. S. Cheng, Y. Wang, Completeness on the stability criterion of fractional order LTI systems, *Fractional Calculus and Applied Analysis* 20 (1) (2017) 159–172.

[16] Y. Q. Chen, Y. H. Wei, X. Zhou, Y. Wang, Stability for nonlinear fractional order systems: an indirect approach, *Nonlinear Dynamics* 89 (2) (2017) 1011–1018.

[17] D. Sheng, Y. H. Wei, S. S. Cheng, J. M. Shuai, Adaptive backstepping control for fractional order systems with input saturation, *Journal of The Franklin Institute* 354 (5) (2017) 2245–2268.

[18] Y. H. Wei, W. T. Peter, Z. Yao, Y. Wang, The output feedback control synthesis for a class of singular fractional order systems, *ISA Transactions* 69 (2017) 1–9.

[19] S. S. Cheng, Y. H. Wei, Y. Q. Chen, Y. Li, Y. Wang, An innovative fractional order LMS based on variable initial value and gradient order, *Signal Processing* 133 (2017) 260–269.

[20] S. S. Cheng, Y. H. Wei, Y. Q. Chen, S. Liang, Y. Wang, A universal modified LMS algorithm with iteration order hybrid switching, *ISA Transactions* 67 (2017) 67–75.

[21] D. Sierociuk, A. Dzieliński, Fractional Kalman filter algorithm for the states, parameters and order of fractional system estimation, *International Journal of Applied Mathematics and Computer Science* 16 (1) (2006) 129–140.

[22] A. Dzielinski, D. Sierociuk, Ultracapacitor modelling and control using discrete fractional order state–space models and fractional Kalman filters, in: *European Control Conference (ECC 2007)*, Kos, 2007, pp. 2916–2922.

[23] A. Kiani-B, K. Fallahi, N. Pariz, H. Leung, A chaotic secure communication scheme using fractional chaotic systems based on an extended fractional Kalman filter, *Communications in Nonlinear Science and Numerical Simulation* 14 (3) (2009) 863–879.

[24] D. Sierociuk, I. Tejado, B. M. Vinagre, Improved fractional Kalman filter and its application to estimation over lossy networks, *Signal Processing* 91 (3) (2011) 542–552.

[25] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, Academic Press, New York, 1982.

[26] M. S. Grewal, A. P. Andrews, *Kalman Filtering: Theory and Practice with MATLAB*, Wiley–IEEE Press, New Jersey.

[27] L. Zhao, *Nonlinear System Filtering Theory*, National Defend Industry Press, Beijing, 2012.