

An innovative parameter estimation for fractional-order systems in the presence of outliers

Rongzhi Cui · Yiheng Wei · Yuquan Chen ·
Songsong Cheng · Yong Wang 

Received: 29 September 2016 / Accepted: 2 March 2017 / Published online: 15 March 2017
© Springer Science+Business Media Dordrecht 2017

Abstract This paper investigates the problem of parameter estimation for fractional-order linear systems when output signal is polluted by noise and outliers. Different from conventional filtering and semi-definite programming methods, the outliers detection problem is formulated as a matrix decomposition problem based on a novel nuclear norm method, which can not only make exact detection of outliers, but also estimate measurement noise at the same time. Then, a new parameter estimation approach is developed via a modified fractional-order gradient method with variable initial value mechanism and fractional-order parameter update law. With the adoption of recovered output signal, the proposed approach can obtain much better estimation performance, whose effectiveness and superiority are verified by strict mathematical analysis and detailed numerical examples.

Keywords Parameter estimation · Outliers detection · Matrix decomposition · Nuclear norm method · Fractional-order gradient · Fractional-order parameter update law

1 Introduction

Fractional-order calculus is an extension of ordinary differentiation and integration to arbitrary order [1]. In recent decades, it has been widely used to describe a complex system more concisely and precisely such as viscoelastic structures [2], heat conduction [3]. It has also attracted increasing attention from the signal processing and systems control community. A tremendous amount of valuable results on system identification [4–6], controllability and observability [7,8], controller synthesis [9,10] of fractional-order systems (FOSs) have been reported.

Outliers occur frequently and cause serious consequences in practice such as signal processing [11,12], mechanical devices [13]. When it comes to system identification, outliers caused by sensor malfunctions and data transmission errors will lead to worse estimation performance. Thus, they should be eliminated from the observed signal. So far, there have been some approaches for outliers detection, such as Hampel filter and robust regression methods. However, the Hampel filter behaves badly with coarsely quantized data [14]. Robust regression methods are inherently less sensitive to outliers, such as the least absolute deviations (LAD)

R. Cui · Y. Wei · Y. Chen · S. Cheng · Y. Wang (✉)
Department of Automation, University of Science and
Technology of China, Hefei 230026, China
e-mail: yongwang@ustc.edu.cn

R. Cui
e-mail: crz@mail.ustc.edu.cn

Y. Wei
e-mail: neudawei@ustc.edu.cn

Y. Chen
e-mail: cyq@mail.ustc.edu.cn

S. Cheng
e-mail: sscheng@mail.ustc.edu.cn

[15]. Recently, Wright et al. put forward robust principal component analysis (RPCA) [16], which decomposed a corrupted matrix into a low-rank matrix and a sparse matrix, and then solved a semi-definite programming problem [17]. However, their algorithms are only for integer systems and need to filter measurement noise before detecting outliers. As we all know, filter is bound to reduce the values of outliers, leading to bad detection result. Unlike above approaches, we formulate the outliers detection problem as a matrix decomposition problem with the aid of a new nuclear norm method, which can detect outliers accurately and estimate noise concurrently. Besides, different from [18, 19], it does not need to construct a noise model in advance. Generally, it is not easy to describe measurement noise using a concrete model. After eliminating the detected outliers and estimated measurement noise from the output signal, we obtain recovered “clean” data for system identification.

Up to now, there have been several kinds of methods for FOSs identification in continuous- and discrete-time domains. Liao et al. [20] studied a subspace algorithm via the Poisson filter. Victor et al. [5] proposed an optimal instrumental variable method for estimating transfer function coefficients. In this paper, we study the parameter estimation problem from a new angle. It is well known that the gradient method is simple and convenient for parametrization model estimation and has attracted wide interest, which is firstly generalized to a fractional one named fractional-order gradient method (FOGM) [21]. Of course, there are other forms and applications of fractional gradient, such as fractional generalization of gradient systems using Riemann–Liouville derivative [22], fractional vector calculus by Grünwald–Letnikov definition [23], while the Caputo’s derivative and its difference are utilized generally in the literature. Though the FOGM algorithm is able to achieve faster convergence speed, it cannot converge to the extreme point [24]. In order to solve this problem, we modify the calculus’s low terminal of FOGM. In addition, it is proved that fractional-order parameter update law can obtain better steady accuracy. Wei et al. extended the conventional gradient estimator to the fractional-order field in continuous case [25, 26]. In the discrete case, the LMS algorithm is extended to fractional-order case, fractional-order parameter update law, by means of replacing the first-order difference with a fractional-order $0 < \alpha < 1$ [27]. It is unfortunate that the contradiction between estimation

accuracy and convergence speed cannot be coordinated taking advantage of both of them. To solve this problem, a novel fractional-order update gradient method (FOUGM) is proposed based on them.

Motivated by above discussions, this paper aims at designing an effective approach to estimate parameters of FOSs when the output signal is disturbed by Gauss white noise and outliers. A novel outliers detection approach is developed via nuclear norm and infinite norm for detecting outliers and estimating noise simultaneously. The modified FOGM with variable initial value mechanism can reduce the effect of non-locality of fractional calculus. With the help of fractional-order parameter update law and the modified FOGM, the proposed algorithm is developed step by step, which significantly improves parameter estimation performance.

After recalling the definitions of widely used fractional-order calculus and a necessary lemma in Sect. 2, Section 3 describes the problem of identifying fractional-order linear system and proposes a novel outliers detection approach. Then the fractional-order update gradient method is presented. Meanwhile, its convergence is also analyzed. In Sect. 4, some numerical simulations are provided to illustrate the validity and superiority of the proposed approach. Conclusions are drawn in Sect. 5.

2 Preliminaries

2.1 Continuous fractional-order calculus

Fractional-order calculus is a generalization and unification of the classical integer-order calculus. In view of its good properties on fractional-order derivative of constant and the initial value of Laplace transformation, the following Caputo’s derivative is adopted in this study

$${}_t_0 \mathcal{D}_t^\alpha f(t) = \frac{1}{\Gamma(m-\alpha)} \int_{t_0}^t \frac{f^{(m)}(\tau)}{(t-\tau)^{\alpha-m+1}} d\tau, \quad (1)$$

where $m - 1 < \alpha < m$, $m \in \mathbb{N}_+$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the Gamma function, and $f(t)$ is a smooth function. To simplify the notation, the fractional derivative of order α with the lower terminal at 0 can be denoted as \mathcal{D}_t^α instead of ${}_0 \mathcal{D}_t^\alpha$.

Alternatively, when $0 < \alpha < 1$, (1) can be rewritten in a form similar as the conventional Taylor series

$${}_t_0 \mathcal{D}_t^\alpha f(t) = \sum_{i=0}^{\infty} \frac{f^{(i+1)}(t_0)}{\Gamma(i+2-\alpha)} (t-t_0)^{i+1-\alpha}. \quad (2)$$

2.2 Discrete fractional-order calculus

Several basic definitions and relevant properties of discrete fractional-order calculus will be presented in this section.

The α th fractional-order sum of a discrete sequence $f(n)$, $n \in \mathbb{N}_0$, is defined as

$$\nabla^{-\alpha} f(n) = \sum_{j=0}^n (-1)^j \binom{-\alpha}{j} f(n-j), \tag{3}$$

where $\alpha > 0$ and $\binom{p}{q} = \frac{\Gamma(p+1)}{\Gamma(q+1)\Gamma(p-q+1)}$.

The conventional m th-order difference is given by

$$\nabla^m f(n) = \sum_{j=0}^m (-1)^j \binom{m}{j} f(n-j). \tag{4}$$

Then the Caputo fractional difference can be defined as

$${}^C\nabla^\alpha f(n) = \nabla^{\alpha-m} \nabla^m f(n), \tag{5}$$

with $m - 1 < \alpha < m$.

The discrete Mittag-Leffler function can be expressed as [28]

$$\mathcal{F}_{\alpha,\beta}(\lambda, n) = \sum_{j=0}^{\infty} \lambda^j \frac{\Gamma(j\alpha + \beta + n - 1)}{\Gamma(j\alpha + \beta)\Gamma(n)}. \tag{6}$$

Before moving on, the following lemma is necessary.

Lemma 1 (See [29]) *If $m - 1 < \alpha < m$, then the solution to the fractional difference system*

$$\nabla^\alpha f(n) = \lambda f(n) + u(n), \lambda \neq 1, \tag{7}$$

with initial conditions $\nabla^k f(n)|_{n=0} = b_k, k = 0, 1, \dots, m - 1$, is unique and is given by

$$f(n) = \sum_{k=0}^{m-1} b_k \mathcal{F}_{\alpha,k+1}(\lambda, n) + \sum_{\tau=1}^n \mathcal{F}_{\alpha,\alpha}(\lambda, \tau) u(n - \tau + 1). \tag{8}$$

The system (7) has the following properties with $u(n) = 0$.

- (i) If $\lambda > 0$ and $\lambda \neq 1$, $f(n)$ is divergent.
- (ii) If $\lambda = 0$, $f(n) = \sum_{k=0}^{m-1} \frac{b_k}{k!} \frac{\Gamma(k+n)}{\Gamma(n)}$, therefore $f(n)$ may be divergent or constant.
- (iii) If $\lambda < 0$, $f(n)$ is convergent. More specifically, if $0 < \alpha \leq 1$, $f(n)$ is monotone convergent for large n . If $1 < \alpha < 2$, $f(n)$ is convergent with or without overshoot.

According to Lemma 1, when $0 < \alpha \leq 1, u(n) = 0$, only $\mathcal{F}_{\alpha,1}(\lambda, n)$ needs to be considered. If $\lambda < 0$ and $n \rightarrow \infty$, the asymptotic line for $\mathcal{F}_{\alpha,1}(\lambda, n)$ is [29,30]

$$\mathcal{F}_{\alpha,1}(\lambda, n) = \sum_{j=0}^{\infty} \lambda^j \frac{\Gamma(j\alpha + n)}{\Gamma(j\alpha + 1)\Gamma(n)} \sim -\frac{1}{\lambda(n-1)^\alpha \Gamma(1-\alpha)}. \tag{9}$$

Under this situation, (8) turns out to be

$$f(n) = b_0 \mathcal{F}_{\alpha,1}(\lambda, n) \sim -\frac{b_0}{\lambda(n-1)^\alpha \Gamma(1-\alpha)}. \tag{10}$$

Moreover, the bigger the $|\lambda|$ or α is, the faster the convergence speed of $f(n)$ is.

3 Parameter estimation of fractional-order linear systems

3.1 Problem description

Consider the fractional-order single-input and single-output plant with the differential equation

$$\sum_{i=0}^n a_i \mathcal{D}^{i\alpha} x(t) = \sum_{j=0}^m b_j \mathcal{D}^{j\alpha} u(t), \tag{11}$$

where $m, n \in \mathbb{N}$ and $m < n$. $u(t)$ and $x(t)$ are the input and output signals, respectively. The commensurate-order α is known and satisfies $0 < \alpha < 1$. a_i ($i = 0, 1, \dots, n - 1$) and b_j ($j = 0, 1, \dots, m$) are constants but unknown. Let $y(t)$ be a noisy observation of $x(t)$ on the interval of t . $\varepsilon(t)$ is the Gauss white noise and $z(t)$ consists of 0 and outliers,

$$y(t) = x(t) + \varepsilon(t) + z(t). \tag{12}$$

Then put (12) into system Eq. (11) and donate $v(t) = \varepsilon(t) + z(t)$, rewriting the equation as

$$\sum_{i=0}^n a_i \mathcal{D}^{i\alpha} y(t) = \sum_{j=0}^m b_j \mathcal{D}^{j\alpha} u(t) + \sum_{i=0}^n a_i \mathcal{D}^{i\alpha} v(t). \tag{13}$$

Unlike the discrete-time case, the difficulty for direct identification method of continuous-time systems lies in inability to measure the differentiation of all kinds of orders of input and output signals, and casting approximate ways on differentiation items enlarges the impacts of noises.

In this paper, the Poisson moment functional method (PMF) [31] is chosen to solve this difficulty. The transfer function of PMF can be given as

$$\mathcal{L}[g_l(t)] = G_l(s) = \left(\frac{\eta}{s + \xi} \right)^l, \tag{14}$$

where s is the Laplace variable and $g_l(t)$ is the l th-order Poisson pulse function. $\xi \in \mathbb{R}^+$ and $\eta \in \mathbb{R}^+$ are, respectively, the Poisson filter constant and gain. More detailed information about PMF can be seen in [6].

The essential idea of pre-filtering by Poisson filter is to find a way to transmit the time derivative in the fractional-order sense to a known and special function. Define the procedure of pre-filtering as follows [6]

$$\begin{aligned} P_{g_l(t)}[\mathcal{D}^\alpha y(t)] &= g_l(t) * \mathcal{D}^\alpha y(t) \\ &= \mathcal{L}^{-1}[G_l(s)s^\alpha Y(s)] \\ &= \mathcal{D}^\alpha g_l(t) * y(t), \end{aligned} \tag{15}$$

where $*$ represents the convolution product operator. Then performing this operation on Eq. (13), one can obtain

$$\begin{aligned} \sum_{i=0}^n a_i P_{g_l(t)}[\mathcal{D}^{i\alpha} y(t)] &= \sum_{j=0}^m b_j P_{g_l(t)}[\mathcal{D}^{j\alpha} u(t)] \\ &+ \sum_{i=0}^n a_i P_{g_l(t)}[\mathcal{D}^{i\alpha} v(t)]. \end{aligned} \tag{16}$$

As the concept of continuous-time noise or outliers is not a trivial extension of the discrete-time case [32], therefore, the data are collected with a sampling period which is so small that the approximate errors are negligible during the numerical computation of fractional derivatives. Moreover, the number of samples is assumed to be large enough to guarantee the convergence of the estimated parameters to the true ones. Prescribe $y(k) = y(kT)$, $u(k) = u(kT)$, $v(k) = v(kT)$, and $y_i(k) = P_{g_l(k)}[\mathcal{D}^{i\alpha} y(k)]$, $u_j(k) = P_{g_l(k)}[\mathcal{D}^{j\alpha} u(k)]$, $V(k) = \sum_{i=0}^n a_i P_{g_l(k)}[\mathcal{D}^{i\alpha} v(k)]$, ($k = 1, 2, \dots, N$). Then, Eq. (16) can be rewritten as

$$\sum_{i=0}^n a_i y_i(k) = \sum_{j=0}^m b_j u_j(k) + V(k). \tag{17}$$

Generally, $a_n = 1$. Eq. (17) is expressed as

$$y_n(k) = \varphi_k^T \theta + V(k), \tag{18}$$

where $\varphi_k = [-y_0(k), -y_1(k), \dots, -y_{n-1}(k), u_0(k), u_1(k), \dots, u_m(k)]^T$ and $\theta = [a_0, a_1, \dots, a_{n-1}, b_0, b_1, \dots, b_m]^T$. The least squares estimate for θ is given by

$$\hat{\theta} = \left[\sum_{k=1}^N \varphi_k \varphi_k^T \right]^{-1} \left[\sum_{k=1}^N \varphi_k y_n(k) \right]. \tag{19}$$

It is obvious that the least squares estimate from (19) will be biased even if only Gauss white noise exists let alone outliers. Because outliers in $y(k)$ may cause big errors in parameter estimation, it should be eliminated from the observed signal.

3.2 Outliers detection and noise estimation

We formulate the outliers detection problem as a matrix decomposition problem with the help of nuclear norm and infinite norm. To our best knowledge, our method is the first one to detect outliers and estimate noise simultaneously.

Construct a Hankel form from $y(k)$ as follows:

$$D = \begin{bmatrix} y(1) & y(2) & y(3) & \cdots & y(q) \\ y(2) & y(3) & y(4) & \cdots & y(q+1) \\ y(3) & y(4) & y(5) & \cdots & y(q+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y(p) & y(p+1) & y(p+2) & \cdots & y(N_D) \end{bmatrix}, \tag{20}$$

where $p + q - 1 = N_D$, $p \leq q \leq N$. Assume the system order is much smaller than the dimension of D . The problem of matrix decomposition is described as

$$D = L + S, \tag{21}$$

where $L = \begin{bmatrix} x(1) & x(2) & \cdots & x(q) \\ x(2) & x(3) & \cdots & x(q+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(p) & x(p+1) & \cdots & x(N_D) \end{bmatrix}$ and $S = \begin{bmatrix} v(1) & v(2) & \cdots & v(q) \\ v(2) & v(3) & \cdots & v(q+1) \\ \vdots & \vdots & \ddots & \vdots \\ v(p) & v(p+1) & \cdots & v(N_D) \end{bmatrix}$.

Our goal is to recover L from (21). In view of integer-order system, L is a low-rank matrix; thus, it can be achieved by solving a rank minimization problem [16] in which S can be described as a sparse matrix. However, here the L is not a low-rank matrix due to the non-locality of fractional calculus. We construct the following convex problem

$$\begin{aligned} \min_v \quad & \|L\|_{2*} + r \|S\|_\infty, \\ \text{s.t.} \quad & L + S = D, \end{aligned} \tag{22}$$

where r is a tuning parameter.

Unlike general rank minimization problem, the rank is replaced by the nuclear norm. Since the fractional-order differential value of a function at some point

mainly depends on recent function values and the effect of earlier function values can be ignored, $y(k)$ is mainly related to recent data. Then the nuclear norm is still a valid criterion. However, because the measurement noise is inevitable in practice, S cannot be described as a sparse matrix, so we replace 0-norm with infinite norm for the first time, enlarging the range of practical application.

Remark 1 One may attempt to filter out the noise before detecting outliers, but the filter will also reduce the values of outliers, causing unsatisfying detection result. By the aid of infinite norm, not only can accurate detection be achieved, but also noise could be estimated to a certain degree at the same time. In addition, exact detection results will always be obtained no matter whether noise exists or not.

3.3 Parameter estimation algorithm

After eliminating the detected outliers and estimated noise, the recovered “clean” data are used to estimate parameters of systems. FOGM is put forward with the analysis of convergence and effectiveness.

Similar to Eq. (18),

$$\bar{y}_n(k) = \bar{\varphi}_k^T \theta + \epsilon(k), \tag{23}$$

where $\bar{\varphi}_k = [-\bar{y}_0(k), -\bar{y}_1(k), \dots, -\bar{y}_{n-1}(k), u_0(k), u_1(k), \dots, u_m(k)]^T$. $\bar{y}_i(k) = P_{g_l(k)}[\mathcal{D}^{i\alpha} \bar{y}(k)]$ with the recovered data $\bar{y}(k)$ from measured output $y(k)$, and $\epsilon(k) = \sum_{i=0}^n a_i P_{g_l(k)}[\mathcal{D}^{i\alpha} \bar{\epsilon}(k)]$, $(k = 1, 2, \dots, N)$. It is worth mentioning that the recovered data $\bar{y}(k)$ have a bit of noise, whereas the values of noise are small. Thanks to the proposed outliers detection method, all outliers can be detected accurately, so $\bar{y}(k)$ does not contain $z(k)$, which is very exciting.

$$\hat{y}_n(k) = \bar{\varphi}_k^T \hat{\theta}_k, \tag{24}$$

where $\hat{\theta}_k$ and $\hat{y}_n(k)$ are the estimations of θ and $\bar{y}_n(k)$, respectively.

Set the prediction error as

$$\begin{aligned} e(k) &= \bar{y}_n(k) - \hat{y}_n(k) \\ &= \bar{\varphi}_k^T (\theta - \hat{\theta}_k) + \epsilon(k) \\ &= \bar{\varphi}_k^T \tilde{\theta}_k + \epsilon(k), \end{aligned} \tag{25}$$

where $\tilde{\theta}_k = (\theta - \hat{\theta}_k)$ is the parameter vector estimation error.

The objective function is chosen as

$$J(\hat{\theta}) = \frac{1}{2} e(k)^2 = \frac{1}{2} [\bar{\varphi}_k^T (\theta - \hat{\theta}_k) + \epsilon(k)]^2. \tag{26}$$

Then the identification of θ in (26) is to find a $\hat{\theta}_k$ that minimizes $J(\hat{\theta})$. The FOGM can be used to derive the recursive equation of $\hat{\theta}_k$, given as [21]

$$\hat{\theta}_{k+1} - \hat{\theta}_k = -\mu_{\hat{\theta}_0} \mathcal{D}_{\hat{\theta}_k}^\gamma J(\hat{\theta}), \tag{27}$$

where $\mu > 0$ is the iteration step size, and $0 < \gamma < 1$. However, it cannot converge to true extreme value [24]. To solve this problem, a modified counterpart is established with the aid of variable initial value mechanism, which is expressed as

$$\hat{\theta}_{k+1} - \hat{\theta}_k = -\mu_{\hat{\theta}_{k-1}} \mathcal{D}_{\hat{\theta}_k}^\gamma J(\hat{\theta}). \tag{28}$$

On the basis of definition (2) and considering the fact that Taylor series cannot be implemented in practice, only the first item is reserved to reduce the complexity, obtaining

$$\hat{\theta}_{k-1} \mathcal{D}_{\hat{\theta}_k}^\gamma J(\hat{\theta}) = \frac{J^{(1)}(\hat{\theta}_{k-1})}{\Gamma(2-\gamma)} (\hat{\theta}_k - \hat{\theta}_{k-1})^{1-\gamma}. \tag{29}$$

The step size of gradient descent method must be positive to guarantee that $\hat{\theta}_k$ is corrected along the negative gradient direction until the extreme value of $J(\hat{\theta})$ is obtained. So we replace $(\hat{\theta}_k - \hat{\theta}_{k-1})^{1-\gamma}$ with $|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}$. Here follows the formula

$$\begin{aligned} \hat{\theta}_{k-1} \mathcal{D}_{\hat{\theta}_k}^\gamma J(\hat{\theta}) &= \frac{J^{(1)}(\hat{\theta}_k)}{\Gamma(2-\gamma)} |\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma} \\ &= -\frac{|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}}{\Gamma(2-\gamma)} \bar{\varphi}_k e(k) \\ &= -\frac{|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}}{\Gamma(2-\gamma)} \bar{\varphi}_k [\bar{\varphi}_k^T (\theta - \hat{\theta}_k) + \epsilon(k)]. \end{aligned} \tag{30}$$

One may find that $J^{(1)}(\hat{\theta}_{k-1})$ is replaced by $J^{(1)}(\hat{\theta}_k)$. This is in order to guarantee the convergence of the algorithm, since $\lim_{k \rightarrow \infty} J^{(1)}(\hat{\theta}_k) |\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma} = 0$ indicates $J^{(1)}(\hat{\theta}_k) = 0$ for $|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma} \neq 0$.

Instead of the first-order difference at the right-hand side of (28), a fractional-order parameter update law is adopted. Combining it with the modified FOGM, the FOGM is proposed, described as

$$\nabla^\beta \hat{\theta}_k = -\mu_{\hat{\theta}_{k-1}} \mathcal{D}_{\hat{\theta}_k}^\gamma J(\hat{\theta}), \tag{31}$$

where $0 < \beta \leq 2$.

According to definition (3), (4) and (5), it is easy to obtain

$$\begin{aligned} \nabla^\beta \hat{\theta}_k &= \nabla^{\beta-m} \hat{\theta}_k \nabla^m \hat{\theta}_k \\ &= \nabla^m \hat{\theta}_k + \sum_{j=1}^k (-1)^j \binom{\beta-m}{j} \nabla^m \hat{\theta}_{k-j} \\ &= \nabla^m \hat{\theta}_k + \sum_{j=0}^{k-1} (-1)^{j+1} \binom{\beta-m}{j+1} \nabla^m \hat{\theta}_{k-j-1}. \end{aligned} \tag{32}$$

Considering $g(k) = (-1)^{k+1} \binom{\beta-m}{k+1}$ and $\hat{\theta}_d(k) = g(k) * \nabla^m \hat{\theta}(k)$, we get

$$\nabla^\beta \hat{\theta}_k = \nabla^m \hat{\theta}_k + \hat{\theta}_d(k-1). \tag{33}$$

Combining (30), (31) and (33), the designed algorithm can be divided into the following two cases.

(i) If $0 < \beta \leq 1, m = 1$.

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \hat{\theta}_d(k-1) + \mu \frac{|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}}{\Gamma(2-\gamma)} \bar{\varphi}_k e(k). \tag{34}$$

(ii) If $1 < \beta \leq 2, m = 2$.

$$\begin{aligned} \hat{\theta}_{k+1} &= 2\hat{\theta}_k - \hat{\theta}_{k-1} - \hat{\theta}_d(k-1) \\ &\quad + \mu \frac{|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}}{\Gamma(2-\gamma)} \bar{\varphi}_k e(k). \end{aligned} \tag{35}$$

Remark 2 As can be seen from formula (34) where is a $\hat{\theta}_d(k-1)$, opposite to integer-order update, it is the weighted sum of historical information of $\hat{\theta}$ from $\hat{\theta}_1$ to $\hat{\theta}_{k-1}$, which is equivalent to making a smooth to parameter. Thus, it can be concluded that fractional-order update law has superiority compared with the integer one. To reduce repetitive work, only the case of $0 < \beta \leq 1$ is considered hereinafter.

The whole process proposed above can be described as follows:

The algorithm of FOGM for fractional-order systems with outliers

- Step 1: Solve the convex problem (22). Then eliminate the detected outliers and estimated noise from the output signal, obtaining recovered output $\bar{y}(k)$
 - Step 2: Pre-filter the input and recovered output data through PMF method described by (15), achieving (24)
 - Step 3: Set initial value of $\hat{\theta}$, just not as zero. Set suitable update order β , differential order γ and step size μ
 - Step 4: Compute $\hat{\theta}$ by calculating (25) and (34) iteratively
-

3.4 Analysis of convergence

As mentioned above $\tilde{\theta}_k = (\theta - \hat{\theta}_k)$, combine (30) and (31). Under Caputo fractional-order calculus definition, $\nabla^\beta \theta = 0$; therefore,

$$\begin{aligned} \nabla^\beta \tilde{\theta}_k &= -\nabla^\beta \hat{\theta}_k \\ &= \mu_{\hat{\theta}_{k-1}} \mathcal{D}_{\hat{\theta}_k}^\gamma J(\hat{\theta}) \\ &= -\mu \frac{|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}}{\Gamma(2-\gamma)} \bar{\varphi}_k e(k) \\ &= -\rho_k \bar{\varphi}_k [\bar{\varphi}_k^T \tilde{\theta}_k + \epsilon(k)], \end{aligned} \tag{36}$$

where $\rho_k = \mu \frac{|\hat{\theta}_k - \hat{\theta}_{k-1}|^{1-\gamma}}{\Gamma(2-\gamma)} > 0$, which is equivalent to variable step size and related to differential order γ .

Since the PMF filter is a linear transformation and $\epsilon(k)$ is Gauss white noise with zero mean, $E[\epsilon(k)] = \sum_{i=0}^n a_i P_{gl(k)} E[\mathcal{D}^{i\alpha} \bar{\epsilon}(k)] = 0$ and $\bar{\varphi}_k$ is not related to $\epsilon(k)$ [6]. Take the expectation of (36) and diagonalize $E[\bar{\varphi}_k \bar{\varphi}_k^T]$ with unitary matrix Q , following

$$\nabla^\beta E[\tilde{\theta}_k] = -\rho_k Q^{-1} \Lambda Q E[\tilde{\theta}_k], \tag{37}$$

or equivalently,

$$\nabla^\beta E[\tilde{\vartheta}_k] = -\rho_k \Lambda E[\tilde{\vartheta}_k], \tag{38}$$

where $E[\tilde{\vartheta}_k] = Q E[\tilde{\theta}_k]$, and $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_{n+m-1}\}$, $\lambda_j > 0, j = 1, 2, \dots, n+m-1$.

Considering (10), the solution to each component of the matrix Eq. (38) is given by

$$\begin{aligned} E[\tilde{\vartheta}_j(k)] &= q_j \mathcal{F}_{\beta,1}(-\rho_k \lambda_j, k) \\ &\sim \frac{q_j}{\rho_k \lambda_j (k-1)^\beta \Gamma(1-\beta)}, \end{aligned} \tag{39}$$

where $q_j = E[\tilde{\vartheta}_j(0)]$ is the initial condition for the j th estimated parameter.

According to Lemma 1, $E[\tilde{\vartheta}_j(k)]$ converges to 0 when $k \rightarrow \infty$. Thus, $E[\hat{\theta}_k]$ converges to θ when k is large enough. Besides, as can be seen from (39), it is easy to get the conclusion that $E[\tilde{\vartheta}_j(k)]$ gets to 0 faster for bigger update order β and step size ρ_k . Equivalently, $E[\hat{\theta}_k]$ converges to θ faster with bigger β and ρ_k .

Remark 3 The adoption of variable initial value mechanism reduces the effect of non-locality of fractional calculus, which enables the FOGM to converge. Thanks to fractional-order parameter update law, the convergence performance of the proposed algorithm is improved significantly. Besides, it can also be applied to filter and other applications.

4 Illustrative examples and analysis

In this section, the effectiveness and superiority of the proposed parameter estimation method are verified by numerical examples. In order to reduce the complexity of calculation, the infinite impulse response is used to approximate the fractional-order sum operator in these examples [33].

4.1 Detect outliers and estimate noise simultaneously

Consider the FOS

$$\mathcal{D}^\alpha x(t) + ax(t) = bu(t), \tag{40}$$

$$y(t) = x(t) + \varepsilon(t) + z(t), \tag{41}$$

where $\varepsilon(t)$ is Gauss white noise with noise level 20 dB and $z(t)$ consists of zeros and outliers. PRBS signal is chosen as the input with sampling time $T = 0.02$ s. Select $\alpha = 0.9, a = 0.5, b = 0.2$. The initial conditions of the FOS are chosen as $x(t) = x(0) = 0, -\infty \leq t < 0$. Because the input signal and Gauss white noise are stochastic, a Monte Carlo experiment is carried out about 20 times. $y(k)$ are the sampling values of $y(t)$. An input signal and corresponding output signal are presented in Fig. 1.

Remark 4 The Grünwald–Letnikov definition is chosen for the numerical implement of this fractional-order system. Besides, other methods could also be acceptable [34].

Case 1 The output signal is polluted only by outliers

The optimization problem (22) is solved by CVX-Toolbox of MATLAB and simulated on a 3.0GHz processor with 4 GB of memory. Let $p = 10, q = 280$ and $r = 0.55$. In order to decrease the complexity and repeatability of computation, data segments of $0 \leq t \leq 5.6$ s, that is to say the number of samples is 280, are chosen. For the sake of comparing our method with that presented in [17] and Hampel filter, the corresponding output signal and detection results are shown in Fig. 2 and Table 1.

As shown in Fig. 2, three methods all can detect outliers accurately when measurement noise does not exist. However, the detection accuracy obtained by the proposed method is prior to that of [17], which is shown in Table 1.

On the other hand, the computation time of Hampel filter, the proposed method and [17] are 1.25, 33.29

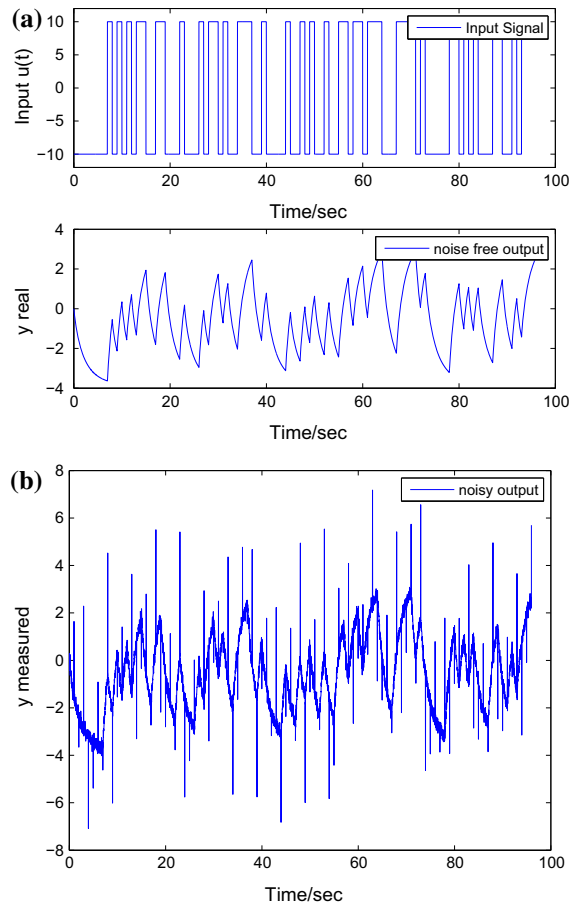


Fig. 1 a Input, noise-free output. b Measured output with white noise and outliers

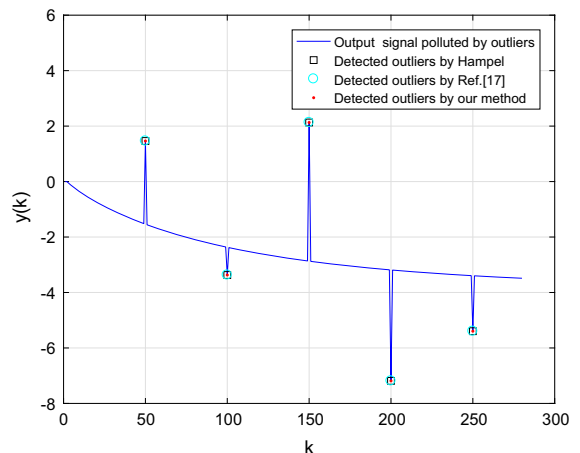


Fig. 2 Comparison of detection results

Table 1 Comparison of outliers detected by [17] and our method

| | $z(1)$ | $z(2)$ | $z(3)$ | $z(4)$ | $z(5)$ |
|------|--------|--------|--------|---------|---------|
| Real | 3 | -1 | 5 | -4 | -2 |
| [17] | 2.9586 | -1 | 4.9237 | -4.0011 | -1.9301 |
| Ours | 2.9952 | -1 | 4.9984 | -4.0001 | -2.0001 |

and 138.14 s, respectively. It is not difficult to interpret why the computation time of [17] is longer than ours. Because its $p = q = \frac{N_D+1}{2}$, which denote the dimension of matrix D , here we select $p = 10, q = N_D + 1 - q$, making the size of D smaller. Thus, the complexity of calculation of the proposed method is between that of other methods. That is to say, the proposed method achieves a trade-off between detection accuracy and computation cost.

Case 2 The output signal is polluted by Gauss white noise and outliers

Since the observation $y(t)$ is often corrupted by measurement noise in practice as well, it is worthwhile to study the problem of outliers detection when $\varepsilon(t) \neq 0$, which is shown in Fig. 1. Note that the total number of sampling $N = 4800$. The output data are divided into ten gropes sequentially, and $p = 10, q = 480$ and $r = 0.55$. Considering the big computation cost of approach presented in [17] and it needs to filter noise firstly, only the Hampel filter is utilized to compare with our method in following simulations.

In Fig. 3a, the number of outliers detected by Hampel filter is 235, while the real number is 90. In other words, some noises are mistakenly determined as outliers by Hampel filter. As shown in Fig. 3b, all the outliers can be detected accurately by the proposed method. Meanwhile, the measurement noise can also be estimated though the estimated values are a bit smaller than the real values. Then the noise level is reduced, which is validated in Fig. 4. Compared with single outliers detection method or filter means, such as RPCA, Hampel filter and LAD, the proposed approach based on nuclear norm and infinite norm possesses effectiveness and superiority, which further enlarges the range of practical application. After eliminating the detected outliers and estimated noise from the output signal, we obtain the recovered output signal.

4.2 Estimate parameters using recovered output data

In following simulations, the recovered output signal is used to estimate parameters of the fractional-order sys-

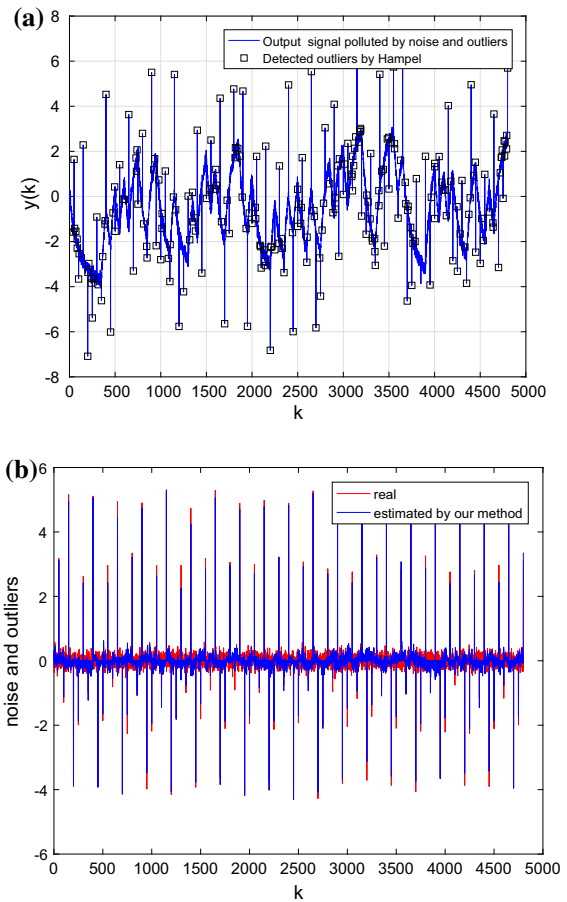


Fig. 3 a Output signal and outliers detected by Hampel filter b Estimated and real noise outliers and counterpart estimated by the proposed method

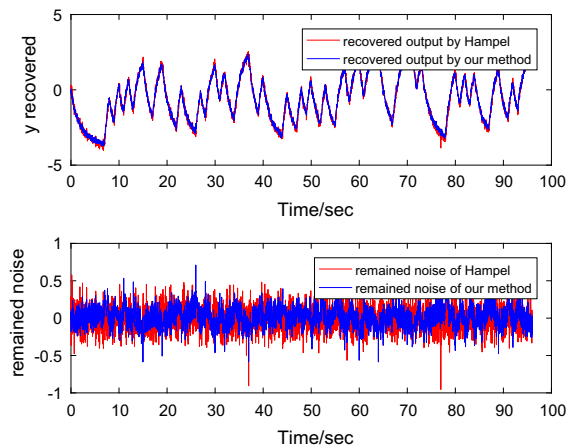


Fig. 4 Recovered output

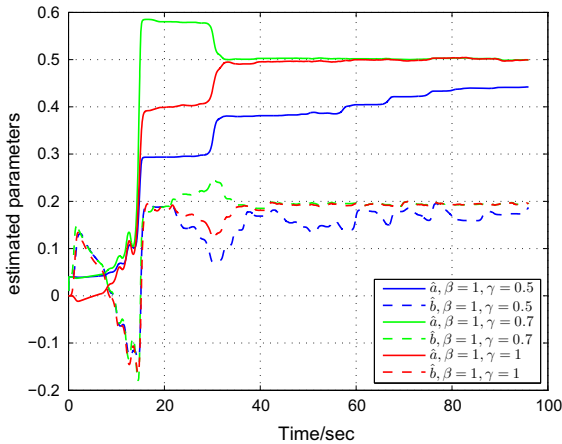


Fig. 5 The effect of γ

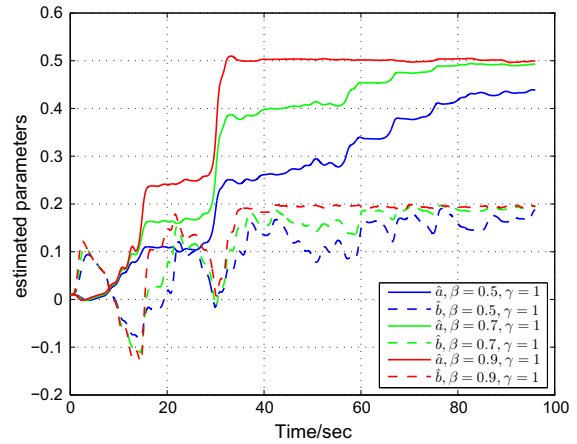


Fig. 6 The effect of β

tem described by Eq. (40). The Poisson filter is selected as $G_l(s) = \left(\frac{0.5}{s+0.5}\right)^2$, in which $\eta = \xi = 0.5, l = 2$, conforming to the conditions of bandwidth and order.

The relative error of parameters estimation is used as the evaluation criterion of the algorithm which is defined as

$$\delta_x = \frac{\|x - x_t\|}{\|x\|}, \tag{42}$$

where x_t is estimation value of the real value x . In the following simulations, data segments of $t \geq 30$ s are selected as steady state that are used to calculate the relative error δ .

Firstly, we consider the effect of differential order γ on estimation performance. In view of the fairness of comparison, let step size $\mu = 0.02$ when $\beta = 1, \gamma = 1$ to achieve the best estimation results of conventional gradient method, achieving the relative error $\delta_a = 1.029\%, \delta_b = 5.740\%$. Set $\mu = 0.25$ when the modified fractional-order gradient method is adopted. Simulation results are shown in Fig. 5.

Note that the convergence speed becomes faster as the differential order γ increases. When $\gamma = 0.7$, the relative error $\delta_a = 0.723\%, \delta_b = 3.609\%$. Comparing cases of $\gamma = 0.7$ and $\gamma = 1$, the estimation accuracy obtained by the modified FOGM is better than that achieved by conventional gradient method.

Secondly, we study the influence of update order β on estimation performance. Set the step size $\mu = 0.04$. As can be seen from simulation results presented in Fig. 6, the bigger the β is, the faster the convergence speed is. When $\beta = 0.9, \gamma = 1$, the relative error $\delta_a = 0.367\%, \delta_b = 2.314\%$.

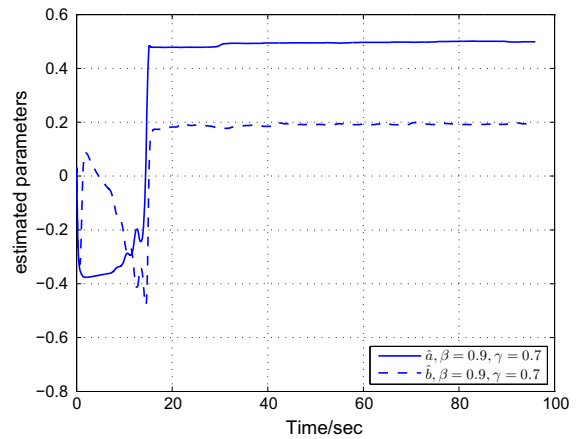


Fig. 7 Simulation results by FOGM

Comparing Fig. 5 with Fig. 6, the modified FOGM is able to achieve faster convergence, while a fractional-order parameter update law could obtain superior estimation accuracy. Then a natural idea is to combine them, getting FOGM, for more satisfactory performance, which is illustrated by a sample simulation example shown in Fig. 7. Here $\gamma = 0.7, \beta = 0.9, \mu = 0.04$.

The relative error $\delta_a = 0.481\%, \delta_b = 2.743\%$. Clearly, faster convergence speed and superior estimation accuracy are gained at the same time. Thus, the contradiction between convergence speed and estimation accuracy is coordinated successfully by the proposed FOGM.

5 Conclusions

In this paper, the FOUGM is proposed for FOSs when measured output is corrupted by Gauss white noise and outliers. Unlike general filter and outliers detection methods, a novel outliers detection approach is developed via nuclear norm and infinite norm. To our best knowledge, it is the first time to exactly detect outliers and estimate noise simultaneously. Then the FOUGM is designed based on a modified FOGM with variable initial value mechanism and fractional-order parameter update law, which can not only obtain faster convergence speed, but also achieve satisfactory estimation accuracy as can be seen from simulation results. Note that the outliers detection method and FOUGM can also be used for other applications and general systems identification, not just for FOSs.

Acknowledgements The authors are grateful to the associate editor and reviewers for their constructive comments based on which the presentation of this paper has been greatly improved. 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

- Machado, J.T., Galhano, A.M., Trujillo, J.J.: On development of fractional calculus during the last fifty years. *Sciencetometrics* **98**(1), 577–582 (2014)
- Wilkie, K., Drapaca, C.S., Sivaloganathan, S.: A nonlinear viscoelastic fractional derivative model of infant hydrocephalus. *Appl. Math. Comput.* **217**(21), 8693–8704 (2011)
- Povstenko, Y.Z.: Fractional heat conduction equation and associated thermal stress. *J. Therm. Stress.* **28**(1), 83–102 (2004)
- Petras, I., Sierociuk, D., Podlubny, I.: Identification of parameters of a half order system. *IEEE Trans. Signal Process.* **60**(10), 5561–5566 (2012)
- Victor, S., Malti, R., Garnier, H., Oustaloup, A.: Parameter and differentiation order estimation in fractional models. *Automatica* **49**(4), 926–935 (2013)
- Hu, Y.S., Fan, Y., Wei, Y.H., Wang, Y., Liang, Q.: Subspace-based continuous-time identification of fractional order systems from non-uniformly sampled data. *Int. J. Syst. Sci.* **47**(1), 122–134 (2016)
- Yan, Z.M.: Controllability of fractional order partial neutral functional integrodifferential inclusions with infinite delay. *J. Frankl. Inst.* **348**(8), 2156–2173 (2011)
- Lu, J.G., Chen, G.R.: Robust stability and stabilization of fractional-order interval systems: an LMI approach. *IEEE Trans. Autom. Control* **54**(6), 1294–1299 (2009)
- Yin, C., Chen, Y.Q., Zhong, S.M.: Fractional order sliding mode based extremum seeking control of a class of nonlinear systems. *Automatica* **50**(12), 3173–3181 (2014)
- Wei, Y.H., Tse, P.W., Yao, Z., Wang, Y.: Adaptive backstepping output feedback control for a class of nonlinear fractional order systems. *Nonlinear Dyn.* **86**(2), 1047–1056 (2016)
- Davy, M., Godsill, S.: Detection of abrupt spectral changes using support vector machines an application to audio signal segmentation. In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 13–17 May 2002, Florida USA, **2**, 1313–1316 (2002)
- Hazel, G.G.: Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **38**(3), 1199–1211 (2000)
- Desforges, M., Jacob, P., Cooper, J.: Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **212**(8), 687–703 (1998)
- Pearson, R.K.: Outliers in process modeling and identification. *IEEE Trans. Control Syst. Technol.* **10**(1), 55–63 (2002)
- Xu, B.C., Liu, X.L.: Identification algorithm based on the approximate least absolute deviation criteria. *Int. J. Autom. Comput.* **9**(5), 501–505 (2012)
- Lu, X.Q., Gong, T., Yan, P., Yuan, Y., Li, X.: Robust alternative minimization for matrix completion. *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* **42**(3), 939–949 (2012)
- Yu, C., Wang, Q.G., Zhang, D., Wang, L., Huang, J.: System identification in presence of outliers. *IEEE Trans. Cybern.* **46**(5), 1202–1216 (2016)
- Smith, R.S.: Frequency domain subspace identification using nuclear norm minimization and Hankel matrix realizations. *IEEE Trans. Autom. Control* **59**(11), 2886–2896 (2014)
- Hu, Y.S., Dai, Y., Liang, Q., Wang, Y.: Subspace-based continuous-time model identification using nuclear norm minimization. *IFAC-PapersOnLine* **48**(28), 338–343 (2015)
- Liao, Z., Zhu, Z.T., Liang, S., Peng, C., Wang, Y.: Subspace identification for fractional order Hammerstein systems based on instrumental variables. *Int. J. Control Autom. Syst.* **10**(5), 947–953 (2012)
- Zahoor, R.M.A., Qureshi, I.M.: A modified least mean square algorithm using fractional derivative and its application to system identification. *Eur. J. Sci. Res.* **35**(1), 14–21 (2009)
- Tarasov, V.E.: Fractional generalization of gradient systems. *Lett. Math. Phys.* **73**(1), 49–58 (2005)
- Meerschaert, M.M., Mortensen, J., Wheatcraft, S.W.: Fractional vector calculus for fractional advection–dispersion. *Phys. A Stat. Mech. Appl.* **367**, 181–190 (2006)
- Pu, Y.F., Zhou, J.L., Zhang, Y., Zhang, N., Huang, G., Siarry, P.: Fractional extreme value adaptive training method: fractional steepest descent approach. *IEEE Trans. Neural Netw. Learn. Syst.* **26**(4), 653–662 (2015)
- Wei, Y.H., Sun, Z.Y., Hu, Y.S., Wang, Y.: On line parameter estimation based on gradient algorithm for fractional order systems. *J. Control Decis.* **2**(4), 219–232 (2015)
- Chen, Y.Q., Wei, Y.H., Liang, S., Wang, Y.: Indirect model reference adaptive control for a class of fractional order sys-

- tems. *Commun. Nonlinear Sci. Numer. Simul.* **39**, 458–471 (2016)
27. Tan, Y., He, Z.Q., Tian, B.Y.: A novel generalization of modified LMS algorithm to fractional order. *IEEE Signal Process. Lett.* **22**(9), 1244–1248 (2015)
 28. Abdeljawad, T., Jarad, F., Baleanu, D.: A semigroup like property for discrete Mittag-Leffler functions. *Adv. Differ. Equ.* **2012**(1), 1 (2012)
 29. Wei, Y.H., Cheng, S.S., Chen, Y.Q., Wang, Y.: Time domain response of nabla discrete fractional order systems. *IEEE Transactions on Automatic Control* (2016) (**Under Review**)
 30. Mainardi, F.: On some properties of the Mittag-Leffler function $E_\alpha(-t^\alpha)$, completely for $t > 0$ with $0 < \alpha < 1$ (2013). ArXiv preprint [arXiv:1305.0161](https://arxiv.org/abs/1305.0161)
 31. Saha, D., Rao, G.P.: A general algorithm for parameter identification in lumped continuous systems—The Poisson moment functional approach. *IEEE Trans. Autom. Control* **27**(1), 223–225 (1982)
 32. Garnier, H., Wang, L., Young, P.C.: Direct Identification of Continuous-time Models from Sampled Data: Issues, Basic Solutions and Relevance. In: Garnier, H., Wang, L., (eds) *Identification of Continuous-time Models from Sampled Data*, pp. 1–29. Springer, London (2008)
 33. Chen, Y.Q., Vinagre, B.M., Podlubny, I.: Continued fraction expansion approaches to discretizing fractional order derivatives an expository review. *Nonlinear Dyn.* **38**(14), 155–170 (2004)
 34. Wei, Y.H., Gao, Q., Peng, C., Wang, Y.: A rational approximate method to fractional order systems. *Int. J. Control Autom. Syst.* **12**(6), 1180–1186 (2014)