Design and analysis of stable differentiator-predictor

Xinhua Wang

Department of Electrical and Electronic Engineering, University of Nottingham,

Nottingham, NG7 2RD, United Kingdom (Email: wangxinhua04@gmail.com)

Abstract: In this paper, the phase lead by the general phase-lead compensator is analyzed, and a criterion of its parameters selection is proposed. In order to get the relatively large time-interval prediction for stochastic signal, two types of differentiator-predictors are presented to estimate future signal and its derivatives: linear high-gain differentiator-predictor and nonlinear differentiator-predictor. Furthermore, a nonlinear extended differentiator-predictor is designed for future estimation and chattering rejection. The stability analysis of differentiator-predictors is described in time domain. The iterative transfer function method is proposed to analyze the robustness in frequency domain, and the rules of differentiator-predictor parameters selection are presented. The analysis shows that the nonlinear differentiator-predictor has stronger adaptability and robustness than the linear high-gain differentiator-predictor. Simulations demonstrate the effectiveness of the proposed methods.

Keywords: Phase-lead compensator, (extended) differentiator-predictor, future signal and its derivatives

1   Introduction

This paper focuses on the problem of prediction of future signal and its derivatives for stochastic signal. Estimation of future signals is important for many applications, for examples, future trajectory prediction for missile interception (Akcal & Ure, 2017), short-term stock prediction (Weng, Lu, Wang, Megahed, & Martinez, 2018), *et al.* Figure 1 shows the future trajectory prediction for missile-interception system. As the speed of attack missile is very fast, the usual methods of real-time interception control need the motion of defense missile much faster and more flexible. In order to get the timely interception, the prediction of future position, velocity and even acceleration of the attack missile is necessary. Moreover, the prediction can relieve the speed-maneuverability requirement for defence missiles. Many times, for prediction, no model information of the attack missile is provided. Therefore, for the interception control system, not only the future position of attack missile needs to be determined, but also its future velocity and acceleration are necessary under condition that the model of attack missile is unknown.

In fact, the implementation of future prediction is to make signal phase in a required lead. The simplest approach for signal prediction is the phase-lead compensator (Franklin, Powell, & Emami-Naeini, 2014; Han, Lee, & Kim, 2018; Kikuuwe, Kanaoka, Kumon, & Yamamoto, 2015): a proper transfer function is used to make the signal phase in lead, and the future prediction can be implemented. However, for the phase-lead compensator: 1) it is only fit for very short time-interval prediction; 2) low accuracy due to large approximations; 3) the derivative of current signal should be known; 4) the method cannot provide the derivatives of future signal; 5) no robustness to high-frequency noise from the frequency characteristic; 6) the parameters selection is not explicit.

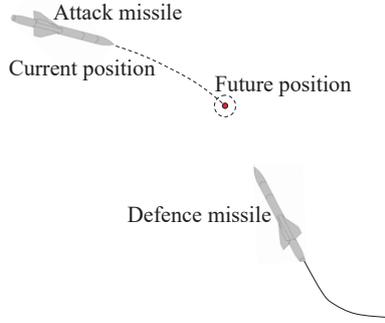For a signal, the Padé approximation can generate a time lag, but not a time lead (Glader,

Figure 1. Future trajectory prediction for missile interception

Hoegnaes, Maekilae, & Toivonen, 1991; Pekař & Kurečková, 2012). If the Padé approximation is used for signal prediction, it will become unstable because a positive pole exists in the system.

Not only the future signal is important, but also the derivatives in the future are necessary for many systems. There are many works on state observers for estimate of unknown states or derivatives (Boukal, Darouach, Zasadzinski, & Radhy, 2017; Hansen, Johansen, Sokolova, & Fossen, 2019), and extended state observers for estimate of system uncertainties (Pu, Yuan, Yi, & Tan, 2015; Sanz, Garcia, Fridman, & Albertos, 2018; Stanković, Rapaić, Manojlović, Mitrović, Simić, & Naumović, 2019). However, for the above observers, the system models are inevitably required. The differentiators can provide the derivatives estimate without considering of system model (Ahrens, & Khalil, 2009; Khalil, 2017; Khalil, & Priess, 2016; Ibrir, 2004; Levant, 1998; Levant, 2003; Levant, & Livne, 2018; Levant & Yu, 2018; Moreno, 2018; Obeid, Fridman, Laghrouche, Harmouche, & Golkani, 2018; Wang, Chen, & Yang, 2007; Wang & Lin, 2012; Wang & Shirinzadeh, 2014). However, the above observers and differentiators can only estimate the current states and derivatives.

Some observers with phase-lead compensation can provide the estimate of future derivatives under condition that the system models are known (Ahmed-Ali, Giri, Krstic, & Kahelras, 2018; Chakrabarty, Fridman, Żak, Buzzard, 2018; Kader, Zheng, & Barbot, 2017).

In this paper, first, a general phase-lead compensator is designed, and the criterion of its parameters selection is proposed. Interestingly, the popular phase-lead compensator (Franklin, Powell, & Emami-Naeini, 2014) is the special case of this general compensator. Second, the phase lead and future derivatives estimation are considered simultaneously, and two types of differentiator-predictors are presented respectively for stochastic signal without considering of system model: linear high-gain differentiator-predictor and nonlinear (extended) differentiator-predictor. The proposed differentiator-predictors have the iterative structure, and each iteration step can estimate the corresponding future signal and its derivatives. At the final step, the future signal and the derivatives in large phase lead will achieve.

In the presented differentiator-predictors, the linear or nonlinear differentiators are adopted as the elements. When the linear differentiator is used in the proposed iterative differentiator-predictor, its gains tend to infinity (Ahrens & Khalil, 2009; Khalil, 2017; Khalil & Priess, 2016; Ibrir, 2004). First, the selection of large gains makes the bandwidth very large, and it is sensitive to high-frequency noise. Second, peaking phenomenon happens. It means that the maximal value of system output during the transient increases infinitely when the gains tend to infinity. Adversely, the peaking

effect becomes worse when an iterative algorithm is used for the linear differentiator-predictor. For the nonlinear differentiators, the gains do not need to be very large (Levant, 1998; Levant, 2003; Levant & Livne, 2018; Levant & Yu, 2018; Moreno, 2018; Obeid, Fridman,Laghrouche, Harmouche, & Golkani, 2018; Wang, Chen, & Yang, 2007; Wang & Shirinzadeh, 2014; Wang & Lin, 2012). Thus, the peaking effect can be avoided. Importantly, through selecting the suitable gains, the bandwidth of the nonlinear differentiator can match that of the predictor in the differentiator-predictor, and its strong robustness can reject the predictor drift with frequency increasing. Although the sliding mode differentiator plays an important role for derivatives estimate (Levant, 1998; Levant, 2003; Levant & Livne, 2018; Levant & Yu, 2018; Moreno, 2018; Obeid, Fridman, Laghrouche, Harmouche, & Golkani, 2018), it is not very fit for the iterative differentiator-predictor design. In fact, the existence of switching function makes the serious chattering happen in the highest-order derivative estimate, although no chattering exists in lower-order derivative estimates. When the high-order sliding mode differentiator is used in an iterative algorithm, the chattering effect will be amplified adversely. Accordingly, the chattering in the iterative differentiator-predictor will contaminate the estimate outputs. The continuous nonlinear differentiators can provide continuous, accurate and smoothed estimations (Wang, Chen, & Yang, 2007; Wang & Shirinzadeh, 2014; Wang & Lin, 2012). In the experiments on helicopter estimation and control (Castañeda, Plestan, Chriette, & León-Morales, 2016), the continuous nonlinear differentiator (Wang & Lin, 2012) was compared with the standard sliding mode differentiator (Levant, 1998). The advantage of the continuous nonlinear differentiator is the reduction of high frequency oscillations with respect to classic discontinuous differentiators, and the system gains are also bounded. Therefore, the continuous structure of nonlinear differentiator with bounded gains is more fit for the iterative algorithm in the presented differentiator-predictor.

In this paper, the iterative transfer function method is presented to analyze the frequency characteristic of the differentiator-predictor. According the input signal information, the differentiator-predictor parameters (including the gains, the prediction time interval and number of iteration steps) are analysed, and a theorem of the parameters selection is proposed.

2 Analysis of future signal prediction

We are interested in designing a signal prediction method to estimate the future signal $f(t + \Delta)$ and its derivatives $f^{(j)}(t + \Delta)$ according to the current stochastic signal $f(t)$, even system model is unknown, and the large future time interval Δ is required. Figure 2 shows that the future estimate and the signal phase leading are equivalent.
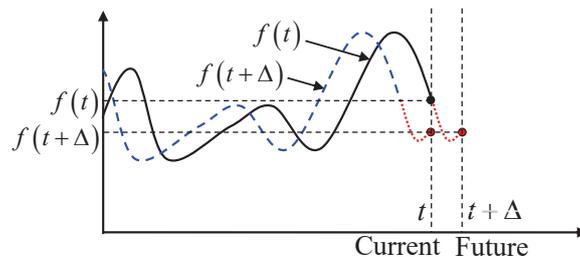

Figure 2. Future prediction scheme

From Figure 2, after phase leading, we can observe $f(t + \Delta)$ at the current time t, where, $f(t + \Delta)$ is the signal value at future time $t + \Delta$. It means that the phase lead can implement the future signal prediction.

In the following, a general phase-lead compensator including the criterion of parameters selection will be presented, and the popular phase-lead compensator (Franklin, Powell, & Emami-Naeini, 2014) is the special case of this general compensator.

2.1 Phase-lead compensator

Lemma 1: Considering a stochastic signal $f(t)$ and a time lead $\Delta \in (0, 1)$, for the phase-lead compensator

$$\frac{Y(s)}{F(s)} = \frac{1 + a_1 s}{1 + a_2 s} \tag{1}$$

if the $a_2 \in (0, 1)$ is selected as small as possible, and the following relations hold:

$$a_1 = a_2 + \Delta \tag{2}$$

$$a_2 < \frac{1}{\max \omega [f(t)]} - \Delta \tag{3}$$

then, there exists a positive constant L, such that

$$\lim_{t \to 0} |y(t) - f(t + \Delta)| \leq L \cdot \Delta \tag{4}$$

where, $f(t)$ is the input signal, and $y(t)$ is the system output; $F(s) = L[f(t)]$ and $Y(s) = L[y(t)]$; $\max \omega [f(t)]$ is the maximal angular frequency of $f(t)$, and $\max \omega [f(t)] < \frac{1}{\Delta}$.

Proof: The phase-lead compensator (1) can be written in time domain by

$$\dot{y}(t) = -\frac{1}{a_2} y(t) + \frac{1}{a_2} f(t) + \frac{a_1}{a_2} \dot{f}(t) \tag{5}$$

For the time lead $\Delta \in (0, 1)$, the reference future signal is $f(t + \Delta)$. The error is defined as

$$e(t) = y(t) - f(t + \Delta) \tag{6}$$

Then, considering of derivative approximation and Taylor expansion, the error system is given by

$$
\begin{aligned}
\dot{e}(t) &= -\frac{1}{a_2} e(t) - \frac{1}{a_2} f(t + \Delta) + \frac{1}{a_2} f(t) + \frac{a_1}{a_2} \dot{f}(t) - \dot{f}(t + \Delta) \\
&= -\frac{1}{a_2} e(t) - \frac{\Delta}{a_2} \frac{(f(t + \Delta) - f(t))}{\Delta} + \frac{a_1}{a_2} \dot{f}(t) - \dot{f}(t + \Delta) \\
&= -\frac{1}{a_2} e(t) - \frac{\Delta}{a_2} \left\{ \dot{f}(t) + \sum_{i=1}^{\infty} \frac{1}{(i+1)!} \Delta^i f^{(i+1)}(t) \right\} + \frac{a_1}{a_2} \dot{f}(t) - \left\{ \dot{f}(t) + \sum_{i=1}^{\infty} \frac{1}{i!} \Delta^i f^{(i+1)}(t) \right\} \\
&= -\frac{1}{a_2} e(t) + \left(\frac{a_1}{a_2} - \frac{\Delta}{a_2} - 1 \right) \dot{f}(t) + O(\Delta)
\end{aligned} \tag{7}
$$

where, $O\left(\Delta\right) = -\frac{\Delta}{a_2} \sum\limits_{i=1}^{\infty} \frac{1}{(i+1)!} \Delta^i f^{(i+1)}\left(t\right) - \sum\limits_{i=1}^{\infty} \frac{1}{i!} \Delta^i f^{(i+1)}\left(t\right)$. In order to make the system stable, and to make the error small sufficiently: i) $\frac{1}{a_2}$ is selected as large as possible, i.e., a_2 is as small as possible; ii) $\Delta \in (0, 1)$ is required to be small enough. Furthermore, to make the second term $\frac{a_1}{a_2} - \frac{\Delta}{a_2} - 1 = 0$ in (7), the parameters a_1 and a_2 satisfy

$$a_1 = a_2 + \Delta \tag{8}$$

Importantly, in order to guarantee more useful elements of $f\left(t\right)$ go through the compensator without distortion, a_1 is selected such that $\frac{1}{a_1} > \max \omega\left[f\left(t\right)\right]$. Then, we can get $a_2 < \frac{1}{\max \omega[f(t)]} - \Delta$. Therefore, the system is stable, and the result (4) holds. This concludes the proof. ∎

Specially, when we define $a_2 = T$ and $a_1 = aT$, the phase-lead compensator (1) can be written by (Franklin, Powell, & Emami-Naeini, 2014)

$$\frac{Y\left(s\right)}{F\left(s\right)} = \frac{1 + aTs}{1 + Ts} \tag{9}$$

where $a > 1$, and $T > 0$ is time constant.

Lemma 2: Considering a stochastic signal $f\left(t\right)$ and a time lead $\Delta \in (0, 1)$, for the phase-lead compensator (9), if $T \in (0, 1)$ is selected as small as possible, and the following relations hold:

$$a = \frac{\Delta}{T} + 1 \tag{10}$$

$$T < \frac{1}{\max \omega\left[f\left(t\right)\right]} - \Delta \tag{11}$$

then, there exists a positive constant L, such that

$$\lim_{t \to 0} \left| y\left(t\right) - f\left(t + \Delta\right) \right| \leq L \cdot \Delta \tag{12}$$

where, $f\left(t\right)$ is the input signal, and $y\left(t\right)$ is the system output; $F\left(s\right) = L\left[f\left(t\right)\right]$ and $Y\left(s\right) = L\left[y\left(t\right)\right]$; $\max \omega\left[f\left(t\right)\right]$ is the maximal angular frequency of $f\left(t\right)$, and $\max \omega\left[f\left(t\right)\right] < \frac{1}{\Delta}$.

Proof: For (1) in Lemma 1, we define

$$a_2 = T, a_1 = aT \tag{13}$$

Then, from $a_1 = a_2 + \Delta$ in Lemma 1 and (13), it follows that $a = \frac{\Delta}{T} + 1$. In order to guarantee more useful elements of $f\left(t\right)$ go through the compensator without distortion, aT selected such that $\frac{1}{aT} > \max \omega\left[f\left(t\right)\right]$. Then, we can get $T < \frac{1}{\max \omega[f(t)]} - \Delta$. Therefore, the system is stable, and the result (12) holds. This concludes the proof. ∎

Example 1: Suppose the time lead $\Delta = 0.2s$, and $\max \omega\left[f\left(t\right)\right] = 4$rad/s. We know that requirement $\max \omega\left[f\left(t\right)\right] = 4 < \frac{1}{\Delta} = 5$ is satisfied. According to $T < \frac{1}{\max \omega[f(t)]} - \Delta = 0.25 - 0.2 = 0.05$, we select $T = 0.02$. Then, $a = \frac{\Delta}{T} + 1 = \frac{0.2}{0.02} + 1 = 11$, and the transfer function of the phase-lead compensator is

$$\frac{Y\left(s\right)}{F\left(s\right)} = \frac{1 + 11 \times 0.02s}{1 + 0.02s} \tag{14}$$

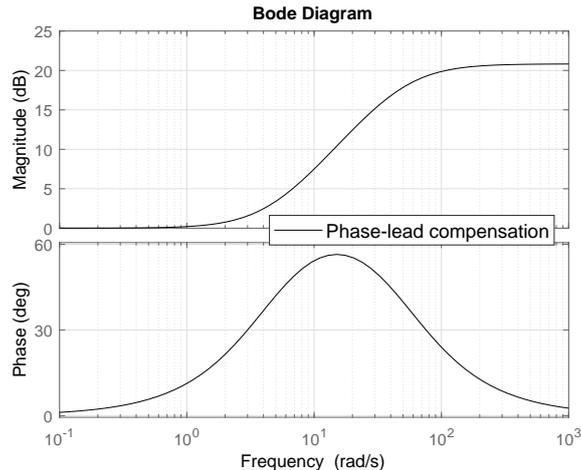The Bode plot of phase-lead compensator (14) is shown in Figure 3.

Figure 3. Bode plot of phase-lead compensator

Although we can get the criterion of parameters selection for the phase-lead compensator, its disadvantages include: 1) it is only fit for small phase lead, i.e., $\Delta \in (0, 1)$; 2) low accuracy due to derivative approximation and low-order Taylor expansion; 3) from (5), it needs the derivative information of current signal, i.e., $\dot{f}(t)$ should be known; 4) it cannot provide the future derivatives; 5) no robustness to high-frequency noise.

2.2 Padé approximation

Can the Padé approximation be used for phase lead? If Padé approximation is used for phase lead, the transfer function can be given by (Glader, Hoegnaes, Maekilae, & Toivonen, 1991; Pekař & Kurečková, 2012)

$$\frac{Y(s)}{F(s)} = e^{s\Delta} \approx \frac{1 + \frac{\Delta}{2}s}{1 - \frac{\Delta}{2}s} \tag{15}$$

where, $Y(s) = e^{s\Delta}F(s)$ i.e., $y(t) = f(t + \Delta)$. However, a positive pole exists in (15), and the system is unstable. Therefore, Padé approximation cannot be used for phase lead. In fact, Padé approximation can only be used for signal lag, i.e., $Y(s) = e^{-s\Delta}F(s)$ in s domain, and $y(t) = f(t - \Delta)$ in time domain. The transfer function of a phase-lag compensator is

$$\frac{Y(s)}{F(s)} = e^{-s\Delta} \approx \frac{1 - \frac{\Delta}{2}s}{1 + \frac{\Delta}{2}s} \tag{16}$$

2.3 Estimation by differentiators

Although the differentiators cannot observe the future signal and its derivatives, they can estimate the current derivatives of signal.

A. Linear high-gain differentiator

Lemma 3 *(Ahrens & Khalil, 2009)*: Considering of a stochastic signal $f(t) = f_0(t) + d(t)$, where, $f_0(t)$ is the desired signal, $d(t)$ is the measurement noise, and $\sup_{t \geq 0} |d(t)| \leq L_d$, for the differentiator:

6

$$\dot{X}(t) = AX(t) + H(f(t) - CX(t)) \tag{17}$$

there exist positive constant c_1 and c_2, such that

$$\lim_{t \to 0} \|F_0(t) - X(t)\| \leq \varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2 \tag{18}$$

and the function $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2$ has the following properties:

1) $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2$ has a global minimum at $\varepsilon = [(n-1)c_2 L_d/c_1]^{(1/n)} \stackrel{\text{def}}{=} c_a L_d^{1/n}$, and

$$\min_{\varepsilon > 0} \left\{ \varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2 \right\} = \left(c_1 c_a + c_2/c_a^{n-1} \right) L_d^{1/n} \stackrel{\text{def}}{=} k_a L_d^{1/n} \tag{19}$$

2) For $\varepsilon \geq c_a L_d^{1/n}$, $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2$ is a strictly increasing function of ε, and $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2 \leq k_b \varepsilon$, where $k_b = c_1 + c_2/c_a^n$;

3) Given $k > 0$, for every $L_d \in [0, (k/k_a)^n)$, there exist $\varepsilon_m = \varepsilon_m(L_d, k) > 0$ and $\varepsilon_M = \varepsilon_M(L_d, k) > c_a L_d^{1/n}$, with $\varepsilon_m \leq c_a L_d^{1/n}$, $\varepsilon_m \leq (L_d c_2 n/k)^{1/(n-1)}$, and $\lim_{L_d \to 0} \varepsilon_M(L_d, k) = k/c_1$, such that $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2 \leq k$ for all $\varepsilon \in (\varepsilon_m, \varepsilon_M]$; where,

$$F_0(t) = \begin{bmatrix} f_0(t) \\ \dot{f}_0(t) \\ \vdots \\ f_0^{(n-1)}(t) \end{bmatrix}, X(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}, H = \begin{bmatrix} \frac{k_n}{\varepsilon} \\ \frac{k_{n-1}}{\varepsilon^2} \\ \vdots \\ \frac{k_1}{\varepsilon^n} \end{bmatrix}, C = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T ; \tag{20}$$

the parameters $k_1, \cdots, k_n$, are selected such that $s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz.

Remark 1: For the linear high-gain differentiator (17), in order to make the system stable, the parameter ε should be selected as small as possible. Thus, the peaking effect is inevitable.

B. Nonlinear differentiator

According to the finite-time stability, the nonlinear differentiator can estimate the current derivatives of signal in spite of the existence of the bounded disturbance.

Lemma 4 *(Wang & Shirinzadeh, 2014)*: For the nonlinear differentiator:

$$\dot{x}_j = x_{j+1} + \frac{k_{n-j+1}}{\varepsilon^j} |f(t) - x_1|^{\alpha_{n-j+1}} \text{sign}(f(t) - x_1); \ j = 1, \cdots, n-1$$

$$\dot{x}_n = \frac{k_1}{\varepsilon^n} |f(t) - x_1|^{\alpha_1} \text{sign}(f(t) - x_1) \tag{21}$$

if the disturbance exists in signal $f(t)$, i.e., $f(t) = f_0(t) + d(t)$, where $f_0(t)$ is the desired signal, $d(t)$ is the bounded stochastic disturbance, and $\sup_{t>0} |d(t)| \leq L_d$, then there exist $\gamma > 1$ and $\Gamma > 0$, such that, for $t \geq \varepsilon \Gamma(\Xi(\varepsilon)e(0))$,

$$\left| x_j - f_0^{(j-1)}(t) \right| \leq L(\delta_{dj})^\gamma, j = 1, \cdots, n \tag{22}$$

7

where, L is the bounded positive constant; $\delta_{dj} = \varepsilon^{n-\frac{j-1}{\gamma}} + \frac{L_d^{\alpha_p}}{h_n}\left(\sum_{i=1}^{n} 2^{1-\alpha_{n-i+1}}k_{n-i+1}\right)\varepsilon^{-\frac{j-1}{\gamma}}$, and

$\delta_{dj} \in (0,1)$, $j = 1,\cdots,n$; $\varepsilon \in (0,1)$; $\sup_{t\geq 0}|f_0^{(n)}(t)| \leq h_n < \infty$; $L_d^{\alpha_p} = L_d^{\alpha_1}$ when $0 < L_d \leq 1$, and $L_d^{\alpha_p} = L_d^{\alpha_n}$ when $L_d > 1$; $k_1, \cdots, k_n$, are selected such that $s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz; $\alpha_1, \cdots, \alpha_n$ satisfy

$$\alpha_{n-j+1} = \frac{j\alpha_1 + n - j}{n}; \; j = 1,\cdots,n \tag{23}$$

with $\alpha_1 \in [0,1)$; $\Xi(\varepsilon) = diag\{1,\varepsilon,\cdots,\varepsilon^{\,n-1}\}$; $e_j = x_j - f_0^{(j-1)}(t)$, $j = 1,\ldots,n$, $n \geq 2$; and $e = [e_1\cdots e_n]^T$.

Lemma 5 *(Properties of error up-boundness for nonlinear differentiator (21)):*

The error up-boundness function $\delta_{dj} = \varepsilon^{n-\frac{j-1}{\gamma}} + \frac{L_d^{\alpha_p}}{h_n}\left(\sum_{i=1}^{n} 2^{1-\alpha_{n-i+1}}k_{n-i+1}\right)\varepsilon^{-\frac{j-1}{\gamma}} \overset{\text{def}}{=} \varepsilon^{n-\frac{j-1}{\gamma}} +$

$cL_d^{\alpha_p}\varepsilon^{-\frac{j-1}{\gamma}}$ in (22) has the following properties:

1) δ_{dj} has a global minimum at $\varepsilon = \left(c\frac{j-1}{n\gamma-j+1}L_d^{\alpha_p}\right)^{(1/n)} \overset{\text{def}}{=} c_a L_d^{\alpha_p/n}$, and

$$\min_{\varepsilon>0}\{\delta_{dj}\} = \left(c_a^{n-\frac{j-1}{\gamma}} + c/c_a^{\frac{j-1}{\gamma}}\right)\left(L_d^{\alpha_p}\right)^{\left(1-\frac{j-1}{n\gamma}\right)} \overset{\text{def}}{=} k_a\left(L_d^{\alpha_p}\right)^{\left(1-\frac{j-1}{n\gamma}\right)} \tag{24}$$

2) For $\varepsilon \geq c_a L_d^{\alpha_p/n}$, δ_{dj} is a strictly increasing function of ε, and $\delta_{dj} \leq k_b\varepsilon^{n-\frac{j-1}{\gamma}}$, where $k_b = (1 + 1/c_a^n)$.

3) For every $L_d \in \left[0, (1/k_a)^{\frac{n\gamma/\alpha_p}{n\gamma-j+1}}\right)$, there exist ε_{*1} and ε_{*2}, such that $\delta_{dj} = \varepsilon^{n-\frac{j-1}{\gamma}} + cL_d^{\alpha_p}\varepsilon^{-\frac{j-1}{\gamma}} < 1$ for all $\varepsilon \in (\varepsilon_{*1}, \varepsilon_{*2})$, where, $0 < \varepsilon_{*1} < \varepsilon_{*2} < 1$.

Proof:

1) From $\frac{\partial(\delta_{dj})}{\partial\varepsilon} = \left(n - \frac{j-1}{\gamma}\right)\varepsilon^{n-\frac{j-1}{\gamma}-1} - \frac{j-1}{\gamma}cL_d^{\alpha_p}\varepsilon^{-\frac{j-1}{\gamma}-1}$, δ_{dj} is strictly decreasing for

$\varepsilon < \left(c\frac{j-1}{n\gamma-j+1}L_d^{\alpha_p}\right)^{(1/n)} \overset{\text{def}}{=} c_a L_d^{\alpha_p/n}$, strictly increasing for $\varepsilon > c_a L_d^{\alpha_p/n}$, and has global minimum at $\varepsilon = c_a L_d^{\alpha_p/n}$ when $\frac{\partial(\delta_{dj})}{\partial\varepsilon} = 0$.

2) We know that $\delta_{dj} = \varepsilon^{n-\frac{j-1}{\gamma}} + c\frac{L_d^{\alpha_p}}{\varepsilon^{\frac{j-1}{\gamma}}} = \left(1 + c\frac{L_d^{\alpha_p}}{\varepsilon^n}\right)\varepsilon^{n-\frac{j-1}{\gamma}}$. When $\varepsilon \geq c_a L_d^{\alpha_p/n}$, $\varepsilon^n \geq c_a^n L_d^{\alpha_p}$.

Therefore, $\delta_{dj} \leq \left(1 + c\frac{L_d^{\alpha_p}}{c_a^n L_d^{\alpha_p}}\right)\varepsilon^{n-\frac{j-1}{\gamma}} = (1 + c/c_a^n)\varepsilon^{n-\frac{j-1}{\gamma}} \overset{\text{def}}{=} k_b\varepsilon^{n-\frac{j-1}{\gamma}}$.

3) For $L_d \in \left[0, (1/k_a)^{\frac{n\gamma/\alpha_p}{n\gamma-j+1}}\right)$, $\min_{\varepsilon>0}\{\delta_{dj}\} = k_a\left(L_d^{\alpha_p}\right)^{\left(1-\frac{j-1}{n\gamma}\right)} < 1$. Because $\varepsilon^{n-\frac{j-1}{\gamma}} + c\frac{L_d^{\alpha_p}}{\varepsilon^{\frac{j-1}{\gamma}}} = 1$

has two solutions at $\varepsilon_{*1} < c_a L_d^{\alpha_p/n}$ and $\varepsilon_{*2} > c_a L_d^{\alpha_p/n}$. Therefore, there exist $0 < \varepsilon_{*1} < \varepsilon_{*2} < 1$, such that $\delta_{dj} = \varepsilon^{n-\frac{j-1}{\gamma}} + c\frac{L_d^{\alpha_p}}{\varepsilon^{\frac{j-1}{\gamma}}} < 1$.

These conclude the the properties 1)-3) ∎.

Remark 2: For the nonlinear differentiator (21), the parameter ε does not need to be selected as small as possible because of the sufficiently large $\gamma > 1$ in the error up-boundness $(\delta_{dj})^\gamma$ and $\delta_{dj} \in (0,1)$. Therefore, no peaking phenomenon happens. Importantly, the differentiator (21) can reduce chattering effect sufficiently because of its continuous structure, and the estimate outputs

are all smooth.

Remark 3: When $\alpha_1 = 0$ is selected, from Theorem 3 in (Wang & Shirinzadeh, 2014), we can get the high-order sliding mode differentiator (Levant, 2003). For this sliding mode differentiator, although high-order sliding mode is introduced, the chattering still exists in the output of the highest-order derivative estimate because of the existence of switching function. If this sliding mode differentiator is used in an iterative algorithm, the chattering will be amplified adversely.

3 Design of differentiator-predictors

Considering stochastic signal $f(t)$, for the given future time interval $\Delta > 0$, two types of differentiator-predictors are designed respectively to estimate future signal and its derivatives $f^{(j-1)}(t + \Delta)$, $j = 1, \cdots, n$.

3.1 High-gain differentiator-predictor

The high-gain technique is used to design an iterative differentiator-predictor to estimate future signal and its derivatives, and one theorem is presented as follows.

Theorem 1: Considering signal $f(t)$, for system

$$\dot{X}_0(t) = AX_0(t) + K(\varepsilon)(f(t) - CX_0(t)) \tag{25}$$

$$\dot{X}_i(t) = AX_i(t) + K(\varepsilon)C\left(e^{A\Delta_i}X_{i-1}(t) - X_i(t)\right); \ i = 1, \cdots, m \tag{26}$$

a constant $L > 0$ exists, such that

$$\lim_{t \to \infty} \left| f^{(j-1)}(t + \Delta) - x_{m,j}(t) \right| \le L\varepsilon; \ j = 1, \cdots, n \tag{27}$$

where, $\Delta > 0$ is the future time interval from the current, and $\Delta = \sum\limits_{i=1}^{m} \Delta_i$, $\Delta_i > 0$, $i = 1, \cdots, m$; m represents the number of iteration steps; the $\varepsilon \in (0, 1)$ is the small parameter;

$$X_0(t) = \begin{bmatrix} x_{0,1}(t) \\ x_{0,2}(t) \\ \vdots \\ x_{0,n}(t) \end{bmatrix}, X_i = \begin{bmatrix} x_{i,1}(t) \\ x_{i,2}(t) \\ \vdots \\ x_{i,n}(t) \end{bmatrix} \tag{28}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}_{n \times n}, C = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T_{n \times 1}, K(\varepsilon) = \begin{bmatrix} \frac{k_n}{\varepsilon} \\ \frac{k_{n-1}}{\varepsilon^2} \\ \vdots \\ \frac{k_1}{\varepsilon^n} \end{bmatrix}_{n \times 1} \tag{29}$$

and $k_1, \cdots, k_n$, are selected such that $s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz.

The proof of Theorem 1 is presented in Appendix.

Remark 4 *(Noise sensitivity and peaking phenomenon):* The high-gain differentiator-predictor provides for the estimations when its gain $1/\varepsilon$ tends to infinity. However, the sensitivity to high-frequency noise increases infinitely. Another drawback of the high-gain differentiator-predictor is its peaking effect: the maximal transient output increases infinitely when $1/\varepsilon$ tends to infinity. Adversely, the peaking becomes more obvious when an iterative algorithm is used.

Remark 5 *(Reduction of peaking phenomenon):* For the high-gain differentiator-predictor (25)-(26), in order to reduce the peaking, the varying ε can be selected as follows:

$$\frac{1}{\varepsilon} = R_0 \left(1 - e^{-\rho \cdot t}\right) \tag{30}$$

or

$$\frac{1}{\varepsilon} = R_0 \frac{1 - e^{-\rho \cdot t}}{1 + e^{-\rho \cdot t}} \tag{31}$$

where $R_0 > 0$; $\rho > 0$ is a small constant. Therefore, the maximal transient output does not grow too large, and peaking effect can be reduced to some extent.

3.2 Continuous nonlinear differentiator-predictor

The finite-time stability of continuous systems is adopted to design an iterative nonlinear differentiator-predictor to estimate the future signal and its derivatives, and one theorem is presented as follows.

Theorem 2: Considering signal $f(t)$, for system

$$\dot{x}_{0,j}(t) = x_{0,j+1}(t) + \frac{k_{n-j+1}}{\varepsilon^j} |f(t) - x_{0,1}(t)|^{\alpha_{n-j+1}} \operatorname{sign}(f(t) - x_{0,1}(t)); \ \ j = 1, \cdots, n-1$$

$$\dot{x}_{0,n}(t) = \frac{k_1}{\varepsilon^n} |f(t) - x_{0,1}(t)|^{\alpha_1} \operatorname{sign}(f(t) - x_{0,1}(t)) \tag{32}$$

$$\dot{x}_{i,j}(t) = x_{i,j+1}(t) + \frac{k_{n-j+1}}{\varepsilon^j} \left|C\left(e^{A\Delta_i} X_{i-1}(t) - X_i(t)\right)\right|^{\alpha_{n-j+1}} \operatorname{sign}\left(C\left(e^{A\Delta_i} X_{i-1}(t) - X_i(t)\right)\right);$$

$$j = 1, \cdots, n-1$$

$$\dot{x}_{i,n}(t) = \frac{k_1}{\varepsilon^n} \left|C\left(e^{A\Delta_i} X_{i-1}(t) - X_i(t)\right)\right|^{\alpha_1} \operatorname{sign}\left(C\left(e^{A\Delta_i} X_{i-1}(t) - X_i(t)\right)\right); \ \ i = 1, \cdots, m \tag{33}$$

there exist a constant $L > 0$, $\varepsilon_* \in (0,1)$, time t_s and $\gamma > 1$, such that, for $\varepsilon \in (\varepsilon_*, 1)$ and $t \geq t_s$

$$\left|f^{(j-1)}(t + \Delta) - x_{m,j}(t)\right| \leq L\varepsilon^{n\gamma - j + 1}; \ \ j = 1, \cdots, n \tag{34}$$

where, $\Delta > 0$ is the future time interval from the current, and $\Delta = \sum_{i=1}^{m} \Delta_i$, $\Delta_i > 0$, $i = 1, \cdots, m$; m is the number of iteration steps; $k_1, \cdots, k_n$, are selected such that $s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz;

$$\alpha_{n-j+1} = \frac{j\alpha_1 + n - j}{n}; \ \ j = 1, \cdots, n \tag{35}$$

with $\alpha_1 \in (0,1)$; and

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}_{n \times n}, C = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T_{n \times 1}, X_{i-1}(t) = \begin{bmatrix} x_{i-1,1}(t) \\ x_{i-1,2}(t) \\ \vdots \\ x_{i-1,n}(t) \end{bmatrix}, X_i(t) = \begin{bmatrix} x_{i,1}(t) \\ x_{i,2}(t) \\ \vdots \\ x_{i,n}(t) \end{bmatrix} \quad (36)$$

The proof of Theorem 2 is presented in Appendix.

Remark 6 *(General form of differentiator-predictors):*

1) For the differentiator-predictor (32)-(33), if $\alpha_1 = 1$ is selected, then high-gain differentiator-predictor (25)-(26) is obtained. Therefore, the differentiator-predictor (32)-(33) includes the high-gain differentiator-predictor (25)-(26).

2) Also, for the differentiator-predictor (32)-(33), when $\alpha_1 = 0$, it becomes the usual high-order sliding mode form. Therefore, the differentiator-predictor (32)-(33) is a general form including linear high-gain differentiator-predictor and the high-order sliding mode differentiator-predictor.

Remark 7 *(Limitation of high-order sliding mode for iterative algorithm):* In fact, although the sliding mode differentiator (Levant, 2003) (when $\alpha_1 = 0$ in (21)) can play an important role for derivatives estimate, it is not very fit for the iterative differentiator-predictor design. The existence of switching function makes the chattering happen in the highest-order derivative estimate, although no chattering exists in lower-order derivative estimates. If the high-order sliding mode is used in an iterative algorithm, the chattering will be amplified adversely, and a vicious circle will happen. The chattering in the iterative differentiator-predictor will change the estimate accuracy adversely. Therefore, the case of $\alpha_1 = 0$ is not preferred for design of iterative differentiator-predictor.

Remark 8 *(No peaking and low chattering for (32)-(33)):* First, the continuous structure of nonlinear differentiator-predictor (32)-(33) can reduce chattering effect sufficiently than the usual high-order sliding mode. Therefore, the estimate outputs are almost smooth. The continuous structure is fit for the iterative algorithm. Second, from Theorem 2, the parameter $\varepsilon \in (\varepsilon_*, 1)$ does not need to be as small as possible. Thus, the peaking effect can be avoided. From Lemma 4, $\gamma > 1$ is sufficiently large, and $n\gamma - j + 1 > 1$ is large enough. Therefore, for any $\varepsilon \in (\varepsilon_*, 1)$, the up-boundness of estimate errors of the differentiator-predictor is small sufficiently.

3.3 Extended nonlinear differentiator-predictor

In order to reduce the chattering effect more effectively, a nonlinear extended differentiator-predictor is presented. In (32)-(33), oscillations exist in the last variable of each-step iteration system. Therefore, we extend the differentiator-predictor (32)-(33) from nth order and m steps into $(n+1)$th order and m steps. For the extended differentiator-predictor, in each step iteration, the first n variables are used for estimate outputs, and the last variable only covers chattering effect. The improved nonlinear differentiator-predictor is presented in the following theorem.

Theorem 3 *(Nonlinear extended differentiator-predictor):* Considering signal $f(t)$, for system

$$\dot{x}_{0,j}(t) = x_{0,j+1}(t) + \frac{k_{n-j+2}}{\varepsilon^j} |f(t) - x_{0,1}(t)|^{\alpha_{n-j+2}} \text{sign}(f(t) - x_{0,1}(t)); \ j = 1, \cdots, n$$

$$\dot{x}_{0,n+1}(t) = \frac{k_1}{\varepsilon^{n+1}} |f(t) - x_{0,1}(t)|^{\alpha_1} \operatorname{sign}(f(t) - x_{0,1}(t)) \tag{37}$$

$$\dot{x}_{i,j}(t) = x_{i,j+1}(t) + \frac{k_{n-j+2}}{\varepsilon^j} \left| C \left(e^{A\Delta_i} X_{i-1}(t) - X_i(t) \right) \right|^{\alpha_{n-j+2}} \operatorname{sign} \left(C \left(e^{A\Delta_i} X_{i-1}(t) - X_i(t) \right) \right);$$
$$j = 1, \cdots, n$$

$$\dot{x}_{i,n+1}(t) = \frac{k_1}{\varepsilon^{n+1}} \left| C \left(e^{A\Delta_i} X_{i-1}(t) - X_i(t) \right) \right|^{\alpha_1} \operatorname{sign} \left(C \left(e^{A\Delta_i} X_{i-1}(t) - X_i(t) \right) \right); \quad i = 1, \cdots, m \tag{38}$$

there exist a constant $L > 0$, $\varepsilon_* \in (0, 1)$, time t_s and $\gamma > 1$, such that, for $\varepsilon \in (\varepsilon_*, 1)$ and $t \geq t_s$

$$\left| f^{(j-1)}(t + \Delta) - x_{m,j}(t) \right| \leq L\varepsilon^{(n+1)\gamma - j + 1}; \quad j = 1, \cdots, n \tag{39}$$

where, $\Delta > 0$ is the future time interval from the current, and $\Delta = \sum_{i=1}^{m} \Delta_i$, $\Delta_i > 0$, $i = 1, \cdots, m$; m is the number of iteration steps; $k_1, \cdots, k_n$ and k_{n+1} are selected such that $s^{n+1} + k_{n+1}s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz;

$$\alpha_{n-j+2} = \frac{j\alpha_1 + n + 1 - j}{n+1}; \quad j = 1, \cdots, n + 1 \tag{40}$$

with $\alpha_1 \in (0, 1)$; and

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}_{n \times n}, C = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}^{T}, X_{i-1}(t) = \begin{bmatrix} x_{i-1,1}(t) \\ x_{i-1,2}(t) \\ \vdots \\ x_{i-1,n}(t) \end{bmatrix}, X_i(t) = \begin{bmatrix} x_{i,1}(t) \\ x_{i,2}(t) \\ \vdots \\ x_{i,n}(t) \end{bmatrix} \tag{41}$$

The proof of Theorem 3 is similar to Theorem 2.

Remark 9: In the extended differentiator-predictor (37)-(38), the variables $x_{i,j}(t)$ (where, $j = 1, \cdots, n$ and $i = 1, \cdots, m$) are used for estimate outputs, and the variables $x_{i,n+1}(t)$ (where, $i = 1, \cdots, m$) only cover the chattering effects. Therefore, for the final-step estimate outputs $x_{m,j}(t)$ (where, $j = 1, \cdots, n$), the chattering effects are reduced sufficiently. Furthermore, it is easy to see that the $(n+1)$th-order differentiator-predictor (37)-(38) provides for a much better accuracy of the future derivatives than the nth-order differentiator-predictor (32)-(33). In fact, $(n+1)\gamma - j + 1$ is larger than $n\gamma - j + 1$. Because $\gamma > 1$ and $\varepsilon \in (\varepsilon_*, 1)$, the estimate error up-boundness in (39) is much smaller than that in (34).

4 Frequency analysis and parameters selection for differentiator-predictor

In this section, we will consider the estimate accuracy, robustness and parameters selection according to the frequency analysis. Furthermore, the determination of the maximal prediction time interval and maximal number of iteration steps will be analysed.

In practice, high-frequency noise exists in signal $f(t)$, and its effect on system outputs is inevitable. Therefore, the robustness to high-frequency noise should be analyzed for the differentiator-predictors. Here, the iterative transfer function method is used to analyze its frequency-domain characteristic. We can find that the differentiator-predictor leads to perform precise estimation and rejection of high-frequency noise to some extent.

4.1 Iterative transfer function of differentiator-predictor

The frequency characteristic of nonlinear differentiator-predictor (32)-(33) is analyzed as follows.

In (32)-(33), for the nonlinear function $|*|^{\alpha_i} sign(*)$, by selecting the input signal as $* = M \sin(\omega t)$, its describing function can be given by:

$$N_j(M) = \frac{2}{M\pi} \int_0^\pi |M \sin(\omega\tau)|^{\alpha_j} sign(M \sin(\omega\tau)) \sin(\omega\tau) d\omega\tau = \frac{\Omega(\alpha_j)}{M^{1-\alpha_j}} \tag{42}$$

where, $\Omega(\alpha_j) = \frac{2}{\pi} \int_0^\pi |\sin(\omega\tau)|^{\alpha_j+1} d\omega\tau$, $\alpha_{n-j+1} = \frac{j\alpha_1+n-j}{n}$, $j = 1, \cdots, n$.

For the differentiator-predictor (32)-(33), define the Laplace transforms $\overline{X}_{i,j}(s) = L[x_{i,j}(t)]$ and $F(s) = L[f(t)]$. Then, we can write (32)-(33) by the form of Laplace transform.

For (32), we get

$$\overline{X}_{0,j}(s) = \frac{s^{j-1} \sum\limits_{q=1}^{n-j+1} \varepsilon^{q-1} \frac{k_q \Omega(\alpha_q)}{M^{1-\alpha_q}} s^{q-1}}{\varepsilon^n s^n + \sum\limits_{i=1}^{n} \varepsilon^{i-1} \frac{k_i \Omega(\alpha_i)}{M^{1-\alpha_i}} s^{i-1}} F(s); \ j = 1, \cdots, n \tag{43}$$

Also, for (33), we get

$$\overline{X}_{i,j}(s) = \frac{s^{j-1} \sum\limits_{q=1}^{n-j+1} \varepsilon^{q-1} \frac{k_q \Omega(\alpha_q)}{M^{1-\alpha_q}} s^{q-1}}{\varepsilon^n s^n + \sum\limits_{i=1}^{n} \varepsilon^{i-1} \frac{k_i \Omega(\alpha_i)}{M^{1-\alpha_i}} s^{i-1}} \overline{X}_{i,e}(s); \ j = 1, \cdots, n \tag{44}$$

where,

$$\overline{X}_{i,e}(s) = \sum\limits_{k=1}^{n} \frac{1}{(k-1)!} \Delta_i^{k-1} \overline{X}_{i-1,k}(s); \ i = 1, \cdots, m \tag{45}$$

According to (43), (44) and (45), the transfer function of the output $x_{m,j}(t)$ to input signal $f(t)$ is determined by

$$\frac{\overline{X}_{m,j}(s)}{F(s)} = a_j(s) \prod\limits_{i=1}^{m} \left(\sum\limits_{k=1}^{n} \frac{1}{(k-1)!} \Delta_i^{k-1} a_k(s) \right) \tag{46}$$

where,

$$a_j(s) = \frac{s^{j-1} \sum\limits_{q=1}^{n-j+1} \varepsilon^{q-1} \frac{k_q \Omega(\alpha_q)}{M^{1-\alpha_q}} s^{q-1}}{\varepsilon^n s^n + \sum\limits_{i=1}^{n} \varepsilon^{i-1} \frac{k_i \Omega(\alpha_i)}{M^{1-\alpha_i}} s^{i-1}}; \ j = 1, \cdots, n \tag{47}$$

4.2 Bode plots of differentiator-predictor with different parameter selections

The effects of the differentiator-predictor parameters on the frequency-domain characteristic are analyzed as follows.

Frequency characteristic with different ε and α_1. With different selections of α_1, we get the values of $\Omega(\alpha_3)$, $\Omega(\alpha_2)$ and $\Omega(\alpha_1)$ shown in Table 1.

Table 1 Values of $\Omega(\alpha_1)$, $\Omega(\alpha_2)$ and $\Omega(\alpha_3)$ with respect to α_1

α_1	$\Omega(\alpha_3)$	$\Omega(\alpha_3)$	$\Omega(\alpha_1)$
1	1	1	1
0.3	1.048	1.104	1.170
0.15	1.060	1.131	1.218

For the differentiator-predictor, the parameters are selected as follows: $n = 3$, $m = 3$; $k_1 = 6$, $k_2 = 11$, $k_3 = 6$; $M = 0.1$; $\varepsilon = 0.9, 0.1, 0.01$; $\alpha_1 = \alpha = 1, 0.3, 0.15$, respectively. When $\alpha_1 = 1$, we can get the high-gain differentiator-predictor (25)-(26). The Bode plots of (46) with different selections of ε and α_1 are described in Figure 4.

From the Bode plot in Figure 4, the differentiator-predictor can obtain the estimation precisely, and the effect of high-frequency noise is reduced to some extent. Also, for the designed differentiator-predictor, from Figure 4, we can find that:

(i) Estimation accuracy: Parameter $\alpha_1 \in (0, 1]$ affects the estimate accuracy: smaller α_1 can obtain more precise estimations; on the other hand, larger α_1 can reduce much noise.

(ii) Bandwidth and robustness: Parameter ε affects the low-pass bandwidth: Decreasing ε, the low-pass bandwidth becomes larger, and the estimate speed increases; increasing ε, the low-pass bandwidth becomes smaller, and much noise is reduced.

4.3 Elements in differentiator-predictor

From the iterative transfer function (46), the differentiator-predictor is made up of the following two elements: iterative predictor and differentiator.

A. Iterative predictor

Before we analyse the performance of the iterative predictor in (46), we introduce the ideal predictor, as follows.

Analysis of ideal predictor (cannot be implemented in practice): For signal $f(t + \Delta)$, the ideal time lead correction is $e^{s\Delta} F(s) = L[f(t + \Delta)]$. The frequency characteristic for the ideal predictor is

$$H(j\omega) = e^{j\omega\Delta} = \cos(\omega\Delta) + j\sin(\omega\Delta) \tag{48}$$

The magnitude-frequency characteristic for the ideal predictor is

$$20\log|H(\omega)| = 20\log\left|e^{j\omega\Delta}\right| = 0 \tag{49}$$

and the phase-frequency characteristic for the ideal predictor is
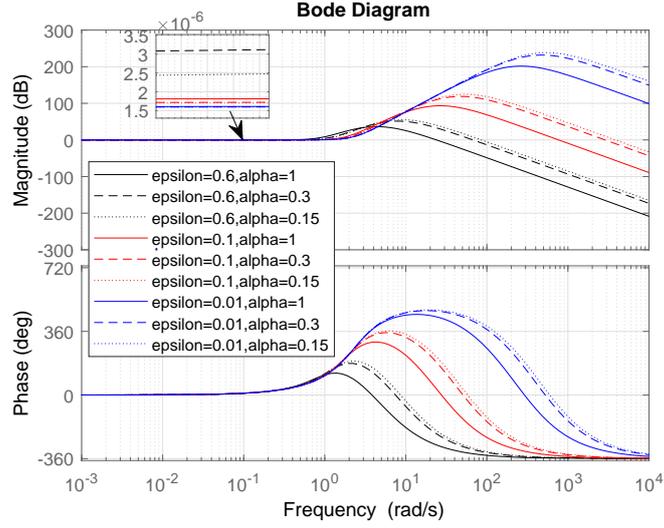
4(a) $\Delta = 1.8s = 0.6s \times 3$ $(\varepsilon = 0.6, 0.1, 0.01; \alpha = 1, 0.3, 0.15)$
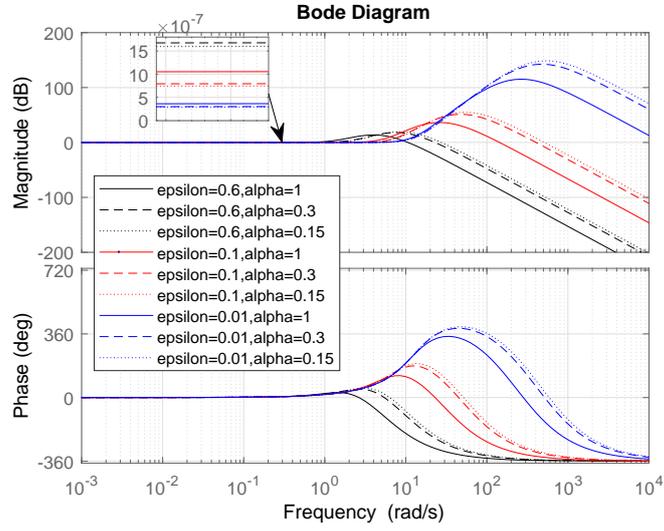
4(b) $\Delta = 0.3s = 0.1s \times 3$ $(\varepsilon = 0.6, 0.1, 0.01; \alpha = 1, 0.3, 0.15)$

Figure 4. Bode plot of predictor-differentiator with different parameter selections

$$\angle H(\omega) = \omega \Delta \tag{50}$$

However, for the above ideal predictor, we cannot implement it in practice.

Conclusion on expression of iterative predictor: The iterative predictor in (46) for signal $f(t)$ can be expressed by

$$e^{s\Delta} F(s) = \left(\prod_{i=1}^{m} \sum_{k=1}^{n} \frac{1}{(k-1)!} \Delta_i^{k-1} s^{k-1} + \prod_{i=1}^{m} O\left(\frac{1}{n!} \Delta_i^n s^n \right) \right) F(s) \tag{51}$$

where, $F(s) = L[f(t)]$; Δ is the prediction time interval; Δ_i is the prediction time interval at the i-th step; $\Delta = \sum_{i=1}^{m} \Delta_i$ and $\Delta_i = \frac{\Delta}{m}$; m is the number of iteration steps.

15

Proof:

$$e^{s\Delta} F(s) = \prod_{i=1}^{m} \left(\sum_{k=1}^{\infty} \frac{1}{(k-1)!} \Delta_i^{k-1} s^{k-1} \right) F(s)$$

$$= \left(\prod_{i=1}^{m} \sum_{k=1}^{n} \frac{1}{(k-1)!} \Delta_i^{k-1} s^{k-1} + \prod_{i=1}^{m} \sum_{k=n+1}^{\infty} \frac{1}{(k-1)!} \Delta_i^{k-1} s^{k-1} \right) F(s)$$

$$\overset{\text{def}}{=} \left(H_m(s) + \prod_{i=1}^{m} O\left(\frac{1}{n!} \Delta_i^n s^n \right) \right) F(s) \tag{52}$$

Therefore, the iterative predictor (51) can implement the time lead correction in time interval $\Delta = \sum_{i=1}^{m} \Delta_i$, where, $\Delta_i \in (0,1)$.

Magnitude-frequency and phase-frequency characteristics for iterative predictor

Here, we analyse the third-order system, i.e., $n = 3$.

1) First, we analyse the case without iteration for the predictor (51), i.e., $m = 1$, and $\Delta_i = \Delta$. The transfer function of predictor can be written by

$$H_1(s) = 1 + \Delta s + \frac{1}{2!} \Delta^2 s^2 \tag{53}$$

Selecting $s = j\omega$, we can get

$$H_1(j\omega) = 1 - \frac{1}{2} \Delta^2 \omega^2 + j\Delta\omega \tag{54}$$

The magnitude of $H_1(j\omega)$ is:

$$|H_1(j\omega)| = \sqrt{\left(1 - \frac{1}{2}\Delta^2\omega^2 \right)^2 + \Delta^2\omega^2} = \sqrt{\left(1 - \frac{1}{2}\left(\frac{\omega}{1/\Delta} \right)^2 \right)^2 + \left(\frac{\omega}{1/\Delta} \right)^2} \tag{55}$$

Therefore, we can get the following magnitude-frequency characteristic:

(i) When $\omega \ll 1/\Delta$, $20\log |H_1(j\omega)| \approx 0$.

(ii) When $\omega = 1/\Delta$, $20\log |H_1(j\omega)| = 20\log \frac{\sqrt{5}}{2} \approx 1\text{dB}$.

(iii) When $\omega \gg 1/\Delta$, $20\log |H_1(j\omega)| \approx 40\log \frac{\Delta\omega}{\sqrt{2}}$.

The argument of $H_1(j\omega)$ is:

$$\angle H_1(j\omega) = \tan^{-1} \frac{\Delta\omega}{1 - \frac{1}{2}\Delta^2\omega^2} = \tan^{-1} \frac{\frac{\omega}{1/\Delta}}{1 - \frac{1}{2}\left(\frac{\omega}{1/\Delta} \right)^2} \tag{56}$$

Therefore, we can get the following phase-frequency characteristic:

(i) When $\omega \ll 1/\Delta$, $\angle H_1(j\omega) \approx \Delta\omega$.

(ii) When $\omega = 1/\Delta$, $\angle H_1(j\omega) = \tan^{-1} \frac{1}{1 - \frac{1}{2}} = \tan^{-1} 2 \approx 28.6°$.

(iii) When $\omega \gg 1/\Delta$, $\angle H_1(j\omega) \approx \tan^{-1} \frac{2}{-\Delta\omega}$.

2) Second, we analyse the multi-step case of iterative predictor (51), i.e., $m > 1$, and $\Delta = \sum_{i=1}^{m} \Delta_i$ and $\Delta_i = \frac{\Delta}{m}$. The transfer function of the predictor can be written by

$$H_m(s) = \prod_{i=1}^{m} \left(1 + \Delta_i s + \frac{1}{2!} \Delta_i^2 s^2 \right) \tag{57}$$

Selecting $s = j\omega$, we can get

$$H_m(j\omega) = \prod_{i=1}^{m} \left(1 - \frac{1}{2} \Delta_i^2 \omega^2 + j \Delta_i \omega \right) \tag{58}$$

The magnitude of $H_m(j\omega)$ is:

$$|H_m(j\omega)| = \prod_{i=1}^{m} \sqrt{ \left(1 - \frac{1}{2} \Delta_i^2 \omega^2 \right)^2 + \Delta_i^2 \omega^2 } = \prod_{i=1}^{m} \sqrt{ \left(1 - \frac{1}{2} \left(\frac{\omega}{1/\Delta_i} \right)^2 \right)^2 + \left(\frac{\omega}{1/\Delta_i} \right)^2 } \tag{59}$$

Therefore, we can get the following magnitude-frequency characteristic:

(i) When $\omega \ll 1/\Delta_i$, $20\log|H_m(j\omega)| \approx 0$.

(ii) When $\omega = 1/\Delta_i$, $20\log|H_m(j\omega)| = 20\log\left(\frac{\sqrt{5}}{2}\right)^m = m20\log\frac{\sqrt{5}}{2} \approx m\mathrm{dB}$.

(iii) When $\omega \gg 1/\Delta_i$, $20\log|H_m(j\omega)| \approx 40\log\left(\frac{\omega\Delta_i}{\sqrt{2}}\right)^m$.

The the argument of $H_m(j\omega)$ is:

$$\angle H_m(j\omega) = \sum_{i=1}^{m} \tan^{-1} \frac{\Delta_i \omega}{1 - \frac{1}{2}\Delta_i^2 \omega^2} = \sum_{i=1}^{m} \tan^{-1} \frac{\frac{\omega}{1/\Delta_i}}{1 - \frac{1}{2}\left(\frac{\omega}{1/\Delta_i}\right)^2} \tag{60}$$

Therefore, we can get the following phase-frequency characteristic:

(i) When $\omega \ll 1/\Delta_i$, $\angle H_m(j\omega) \approx m\Delta_i \omega = \Delta\omega$.

(ii) When $\omega = 1/\Delta_i$, $\angle H_m(j\omega) = m\tan^{-1}\frac{1}{1-\frac{1}{2}} = m\tan^{-1}2$.

(iii) When $\omega \gg 1/\Delta_i$, $\angle H_m(j\omega) \approx m\tan^{-1}\frac{2}{-\Delta_i\omega}$.

3) Comparison of the single step and multi-step predictors:

(i) They have the same prediction time interval Δ. In fact, $\angle|H_1(j\omega)| = \angle|H_m(j\omega)| \approx m\Delta_i\omega = \Delta\omega$.

(ii) The natural angular frequency of the iterative predictor (57) is greater than that of single step predictor (53). In fact, $\omega_1 = 1/\Delta$ for the single step predictor, and $\omega_m = 1/\Delta_i$ for the multi-step predictor. We know that $\omega_m = 1/\Delta_i \gg \omega_1 = 1/\Delta_i$. Therefore, the effective bandwidth $[0, \omega_m]$ of iterative predictor (57) is wider than the $[0, \omega_1]$ of single step predictor (53).

Therefore,

(iii) the iterative predictor (57) is more fit for the input signals with wider bandwidth: slow time-varying and fast time-varying signals.

In the following, we use an example to demonstrate: the effective bandwidth of the iterative predictor is wider than that of the single step predictor.

17

Example 2 *(iterative predictor):* Suppose the prediction time interval $\Delta = 0.6s$.

1) For the single step predictor ($m = 1$ and $\Delta = 0.6s$), the natural frequency $\omega_1 = 1/\Delta = 1/0.6 = 1.67$rad/s, and at this frequency, $20\log |H_1(j\omega)| = 20\log \frac{\sqrt{5}}{2} = 1$dB.

2) For the iterative predictor ($m = 3$ and $\Delta = \Delta_i \cdot m = 0.2s \times 3$), the natural frequency $\omega_m = 1/\Delta_i = 1/0.2 = 5$rad/s, and $20\log |H_m(j\omega)| = 20\log \left(\frac{\sqrt{5}}{2}\right)^3 = 3$dB.

3) We know that $\omega_m > \omega_1$, and the effective bandwidth of iterative predictor is wider than that of single step predictor. However, for the single step predictor, at the frequency $\omega = 1/\Delta_i = 1/0.2 = 5$rad/s, we can get the magnitude $20\log |H_1(j\omega)| = 20\log \sqrt{\left(1 - \frac{1}{2}\Delta^2\omega^2\right)^2 + \Delta^2\omega^2} = 20\log \frac{\sqrt{85}}{2} = 13$dB.

4) Comparison. At the frequency $\omega = 1/\Delta_i = 5$rad/s, the magnitude by the iterative predictor is 3dB, and the magnitude by the single step predictor is 13dB. Therefore, the effective bandwidth of the iterative predictor is wider than that of the single step predictor. Under the same prediction time interval, the output accuracy by the iterative predictor is much higher than that of the single step predictor.

We use the Bode plots to show the frequency characteristic of the single-step and iterative predictors, respectively: the parameters are selected as: $\Delta = 0.1, 0.2, 0.6s$ (where, $m = 1$), and $\Delta = 0.2s \times 3 = 0.6s$ (where, $m = 3$). The Bode plots are shown in Figure 5. From the Bode plots in Figure 5, with frequency increasing, the drift starts to happen at frequency $\omega_1 = 1.67$ for the single step predictor. However, the drift does not happen until $\omega_m = 5$rad/s for the iterative predictor.
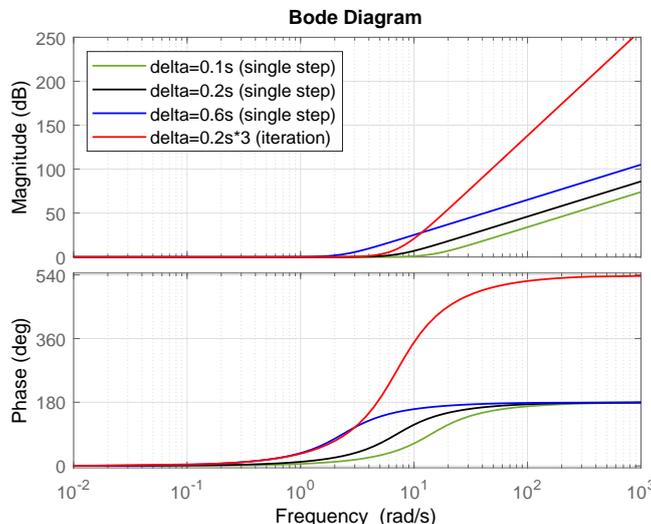
Figure 5. Bode plot of predictor with different prediction time intervals

Remark 10 *(Predictor drift):* From the Bode plots of the predictors in Figure 5, we can find, with frequency increasing, the drift happens when $\omega > \omega_1 = 1/\Delta$ for the single step predictor, and it happens when $\omega > \omega_m = 1/\Delta_m$ for the iterative predictor. The following iterative differentiator can be used to reject the predictor drift, and to provide the derivatives information for the predictor.

B. Iterative differentiator (Estimation, robustness and predictor drift correction)

18

The iterative differentiator provides the derivatives at each step, and it can reject the drift of predictor.

From (46), we can get the transfer function of iterative differentiator, as follows:

$$a_j\left(s\right) = \frac{s^{j-1}\sum\limits_{q=1}^{n-j+1}\varepsilon^{q-1}\frac{k_q\Omega(\alpha_q)}{M^{1-\alpha_q}}s^{q-1}}{\varepsilon^n s^n + \sum\limits_{i=1}^{n}\varepsilon^{i-1}\frac{k_i\Omega(\alpha_i)}{M^{1-\alpha_i}}s^{i-1}};\ \ j = 1,\cdots,n \tag{61}$$

We use the Bode plot to explain its derivative estimations and robustness of the differentiator.

The parameters of differentiator are selected as: $n = 3$, $m = 3$; $k_1 = 6$, $k_2 = 11$, $k_3 = 6$; $M = 0.1$; $\varepsilon = 0.9, 0.1, 0.01$; $\alpha_1 = \alpha = 1, 0.3, 0.15$, respectively. The Bode plots of signal estimate, the first-order and second-order derivatives estimate are shown in Figure 6. We can find that the high-frequency noise can be reduced sufficiently. Parameter ε affects the low-pass frequency bandwidth: with smaller $\varepsilon \in (0,1)$, the low-pass frequency bandwidth is larger; with larger $\varepsilon \in (0,1)$, the low-pass frequency bandwidth is smaller, and much noise can be reduced.

Predictor drift correction by the iterative differentiator: Considering of the combination of the iterative differentiator and predictor, the Bode plot in Figure 7 shows that the differentiator can reject the drift from the predictor (Also, see Figure 4). It means that the combination of differentiator and predictor, i.e., the differentiator-predictor, is stable with strong robustness. Not only the differentiator can reject the predictor drift, but also the high-frequency noise is filtered sufficiently.

Remark 11 *(Comparison of linear and nonlinear differentiators):* For the linear differentiator, the parameter ε is selected as small as possible. One the one hand, peaking effect happens. One the other hand, the selection of small ε makes the bandwidth very large, and it is sensitive to high frequency noise. Also it cannot reject the predictor drift effectively (See Figure 4). For the nonlinear differentiator, the parameter ε does not need to be very small. Thus, the peaking effect can be avoided. Importantly, through regulating the parameter ε, the bandwidth of the nonlinear differentiator is fit for the predictor, and its strong robustness can reject the predictor drift with frequency increasing. From the Bode plots in Figure 4, small overshoot happens when $\varepsilon = 0.6$, but large overshoot happens when $\varepsilon = 0.1, 0.01$, respectively.

4.4 Differentiator-predictor performance analysis:

1) Future signal and its derivatives estimation: For the differentiator-predictor, $x_{m,j}$ estimates $f^{(j-1)}\left(t + \Delta\right)$, respectively, where, $j = 1, \cdots, n$. In the estimate error (34), due to $\varepsilon \in (0,1)$ and $n\gamma - j + 1 \gg 1$, the error $O\left(\varepsilon^{n\gamma-j+1}\right)$ is sufficiently small.

2) Estimate accuracy and robustness by parameter $\varepsilon \in (0,1)$: Decreasing ε, the low-pass bandwidth becomes larger, and the estimate speed increases; increasing ε, the low-pass bandwidth becomes smaller, and much noise is reduced.

3) Estimate accuracy and robustness by parameter $\alpha_1 \in (0,1]$: From Figure 4, smaller α_1 can obtain more accurate estimations; on the other hand, larger α_1 can reduce much noise.

4) Predictor drift correction: The iterative nonlinear differentiator can reject the predictor drift with frequency increasing, and it can provide the derivatives information for the predictor.
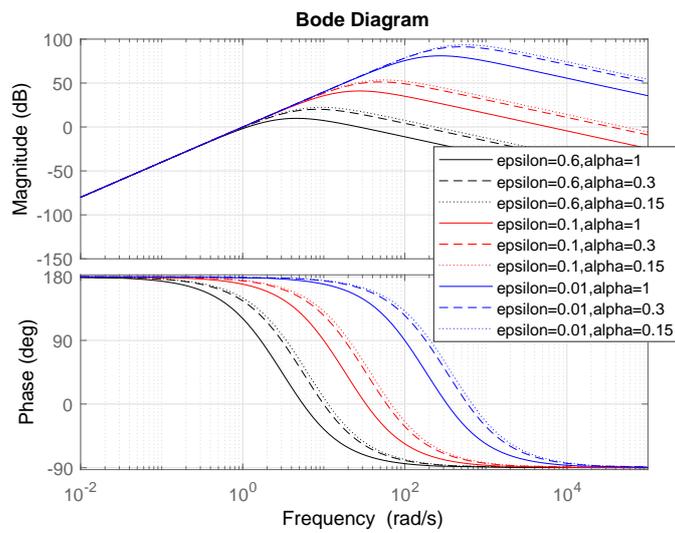
5) Bandwidth affected by Δ_i in the prediction time interval Δ (where, $\Delta_i = \frac{\Delta}{m}$): Relatively larger
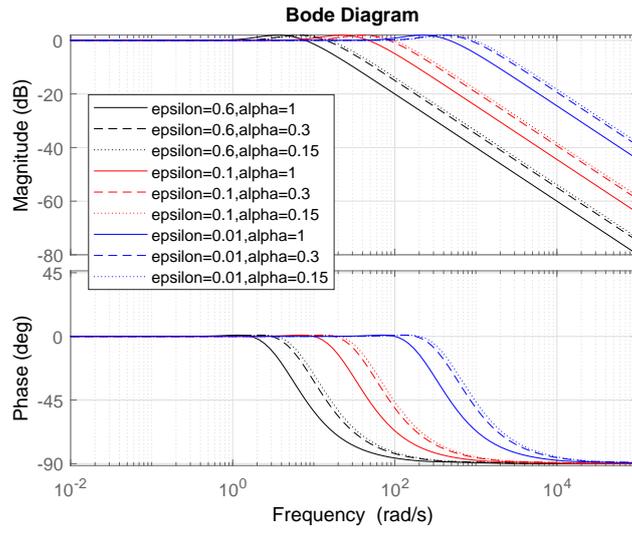
6(a) Signal estimate
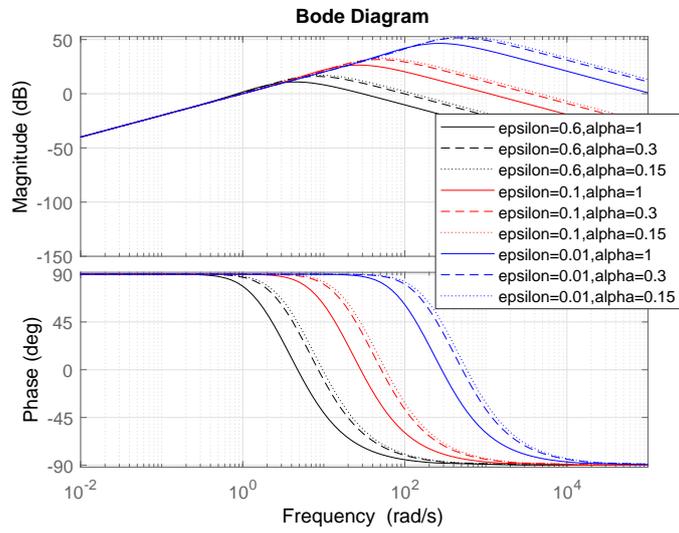
6(b) First-order derivative estimate

6(c) Second-order derivative estimate

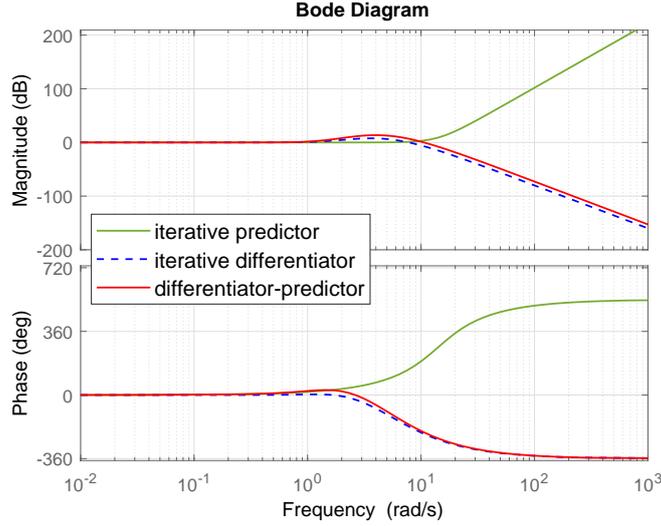Figure 6. Frequency characteristic of differentiator

20

Figure 7. Differentiator corrects predictor drift

Δ_i makes the effective bandwidth smaller; relatively smaller Δ_i makes the effective bandwidth larger (See the Bode plot in Figure 5).

6) Prediction time interval Δ affected by input signal: In Figure 5, from the Bode plot of predictor with different prediction time intervals, we can get:

(i) For fast time-varying (relatively high frequency) input signal, relatively small time interval can be predicted;

(ii) For slow time-varying (relatively low frequency) input signal, relatively large time interval can be predicted.

4.5 Rules of differentiator-predictor parameters selection:

1) Stability conditions: The parameters $k_1, \cdots, k_n$ are selected such that $s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz; and $\alpha_1, \cdots, \alpha_n$ satisfy (35).

2) Estimate accuracy: According to the requirements for differentiator, select ε and α_1, such that: (i) The selection of ε satisfies that the bandwidth of the differentiator is same to that of the predictor. (ii) $\alpha_1 \in (0, 1)$ guarantees the system is continuous; and α_1 decreases, we can get more accurate estimation.

3) Filtering: If much noise exists, $\varepsilon \in (0, 1)$ should increase, or $\alpha_1 \in (0, 1)$ increases, to make the low-pass bandwidth narrow (See Figure 4).

4) Selection of prediction time interval Δ and iteration steps number m

Theorem 4: (i) The prediction time interval at each step satisfies:

$$\Delta_i \leq \frac{1}{\omega_f} \tag{62}$$

where, $\max \omega [f(t)]$ is the maximal angular frequency of input signal $f(t)$, and there exists a constant $\omega_f > 0$, such that $\max \omega [f(t)] \leq \omega_f$.

(ii) The number of iteration steps:

21

$$m \leq \frac{a_N}{20 \log \frac{\sqrt{5}}{2}} = 1.032 a_N \ , \ \text{or} \ \max\{m\} = int \left\{ \frac{a_N}{20 \log \frac{\sqrt{5}}{2}} \right\} \tag{63}$$

where, a_N is the maximal magnitude-frequency characteristic in decibel at the natural frequency, i.e., $\max \left[20 \log |H_m (j\omega)| \right] = a_N$dB.

(iii) The prediction time interval:

$$\Delta = \Delta_i \cdot m \tag{64}$$

(iv) The maximal prediction time interval:

$$\max\{\Delta\} = \frac{1}{\max \omega \left[f(t) \right]} int \left\{ \frac{a_N}{20 \log \frac{\sqrt{5}}{2}} \right\} \tag{65}$$

Proof of Theorem 4:

(i) Suppose we can determine that the angular frequency of input signal $f(t)$ is bounded, and $\max \omega \left[f(t) \right] \leq \omega_f$ holds. In order to make all the useful elements of signal go through the differentiator-predictor, we need to select Δ_i to make

$$\omega_f \leq \omega_m = \frac{1}{\Delta_i} \tag{66}$$

where, $\omega_m = \frac{1}{\Delta_i}$ is the system natural frequency. Therefore, we can get

$$\Delta_i \leq \frac{1}{\omega_f} \tag{67}$$

(ii) We know that $20 \log |H_m (j\omega)| = 20 \log \left(\frac{\sqrt{5}}{2} \right)^m$ at the natural frequency $\omega_m = \frac{1}{\Delta_i}$. Giving a maximal magnitude-frequency characteristic a_NdB at the natural frequency which the system can accept normally, we can get

$$\max \left[20 \log |H_m (j\omega)| \right] = \max \left[20 \log \left(\frac{\sqrt{5}}{2} \right)^m \right] = a_N \tag{68}$$

Therefore, we get

$$20 \log \left(\frac{\sqrt{5}}{2} \right)^m \leq a_N \tag{69}$$

Then, we get the iteration number

$$m \leq \frac{a_N}{20 \log \frac{\sqrt{5}}{2}} = 1.032 a_N \ , \ \text{or} \ \max\{m\} = int \left\{ \frac{a_N}{20 \log \frac{\sqrt{5}}{2}} \right\} \tag{70}$$

(iii) By considering of each-step prediction time interval Δ_i and the iteration steps number m, we can get the prediction time interval

22

$$\Delta = \Delta_i \cdot m \tag{71}$$

(iv) From $\max \omega\left[f\left(t\right)\right] \le \omega_f$, (67) and (70), we can get the maximal value of Δ_i, i.e., $\max\left\{\Delta_i\right\} = \frac{1}{\max \omega[f(t)]}$, and the maximal value of iteration steps number $\max\left\{m\right\} = int\left\{\frac{a_N}{20\log\frac{\sqrt{5}}{2}}\right\}$. Therefore, the maximal prediction time interval is

$$\max\left\{\Delta\right\} = \max\left\{\Delta_i\right\}\max\left\{m\right\} = \frac{1}{\max \omega\left[f\left(t\right)\right]} int\left\{\frac{a_N}{20\log\frac{\sqrt{5}}{2}}\right\} \tag{72}$$

We found that the maximal prediction time interval is decided by the frequency band of input signal and the requirement of estimate accuracy.

This concludes the proof. ∎

Example 3 *(Determination of Δ and m)*: Suppose the maximal angular frequency of a signal $f\left(t\right)$ is 3rad/s, i.e., $\max \omega\left[f\left(t\right)\right] = 3\text{rad/s}$. Suppose we can know $\max \omega\left[f\left(t\right)\right] \le 4\text{rad/s}$. Therefore, according to (67), the prediction time interval at each step should satisfy $\Delta_i \le \frac{1}{4} = 0.25s$. We can select $\Delta_i = 0.2s$. We require the maximal magnitude at the natural frequency $\omega_m = \frac{1}{\Delta_i}$ is $a_N = 3\text{dB}$. Then, from (70), we can get the iteration steps number $\max\left\{m\right\} = int\left\{\frac{3}{20\log\frac{\sqrt{5}}{2}}\right\} = 3$. Therefore, from (71), the the prediction time interval of the differentiator-predictor can be selected as $\Delta = \Delta_i \cdot m = 0.2s \times 3 = 0.6s$. Furthermore, according to (72), the maximal prediction time interval is $\max\left\{\Delta\right\} = \frac{1}{\max \omega[f(t)]} int\left\{\frac{a_N}{20\log\frac{\sqrt{5}}{2}}\right\} = \frac{3}{3} = 1s$.

5 Simulations

In this section, we use the simulations to demonstrate the presented differentiator-predictors by considering of different types of input signals: constant signal, step signal, ramp signal, and parabola. Furthermore, slow time-varying and fast time-varying signals are used to illustrate the relation between input signal band and the prediction time interval.

5.1 The following time functions are selected as the input signal $f(t)$, respectively:

(i) Constant signal: $f\left(t\right) = 1$; (ii) Step signal: $f\left(t\right) = \begin{cases} 0, \text{ where, } t < 1.5s \\ 1, \text{ where, } t \ge 1.5s \end{cases}$; (iii) Ramp signal: $f\left(t\right) = 10t$; (iv) Parabola: $f\left(t\right) = 0.5t^2$

The future time interval $\Delta = 1.5s$ is considered, and $\Delta = \sum\limits_{i=1}^{3} \Delta_i = 0.5s \times 3 = 1.5s$ is assigned. The nonlinear differentiator-predictor (32)-(33) is used for the estimation, and the parameters are: $n = 3$, $m = 3$; $k_1 = 6$, $k_2 = 11$, $k_3 = 6$; $1/\varepsilon = 1.1$; $\alpha_1 = 0.2$.
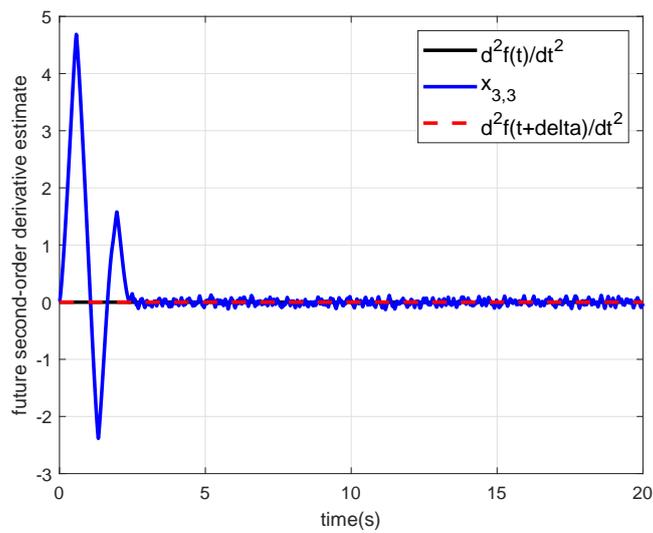
Figures 8-11 show the future derivatives estimation by the nonlinear differentiator-predictor for the different types of input signals, respectively. The precise prediction and estimation are implemented simultaneously.
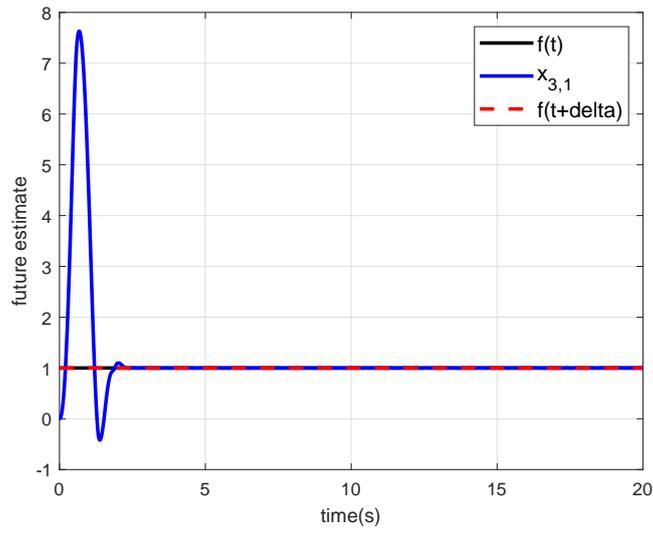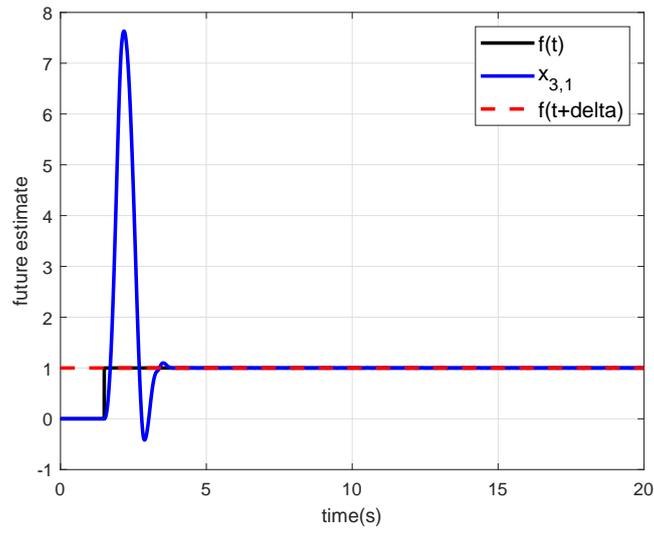
8(a) Future signal estimate
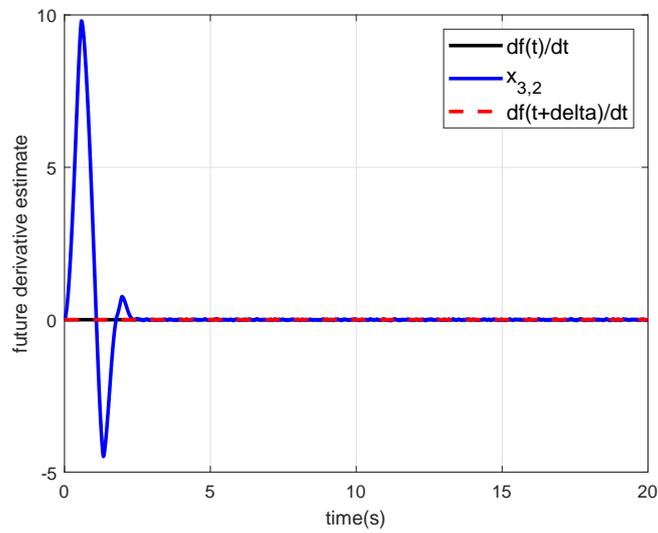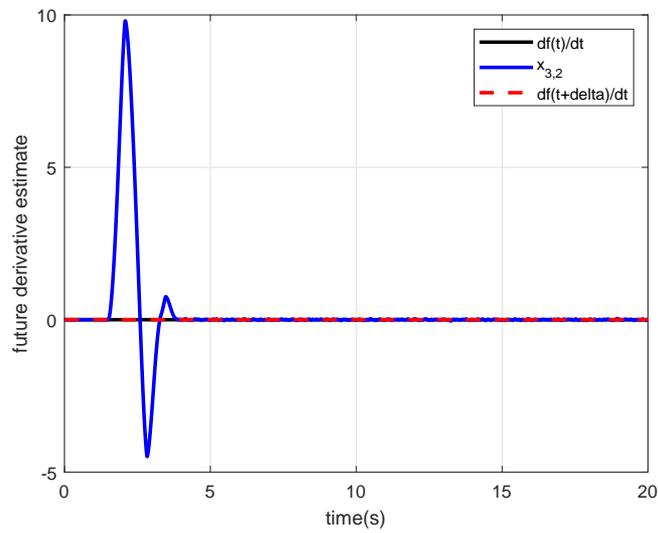

8(b) Future derivative estimate

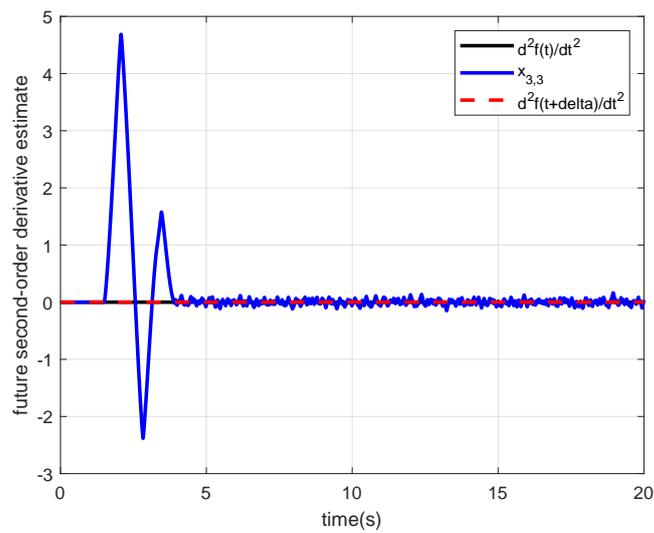8(c) Future second-order derivative estimate

Figure 8. Nonlinear differentiator-predictor with constant input signal
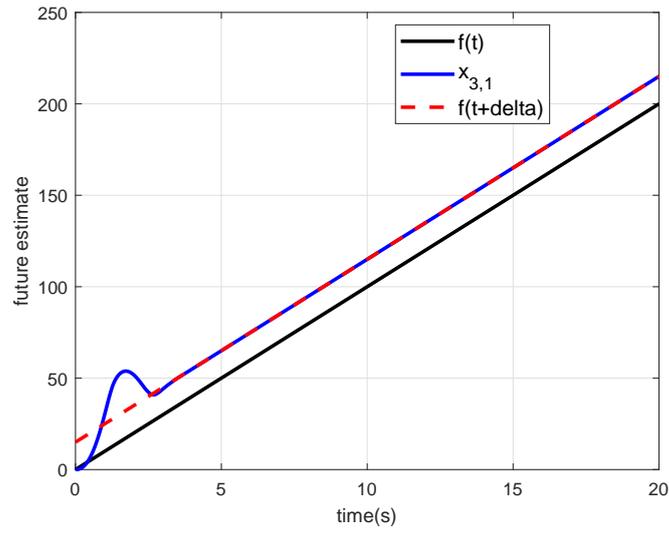
9(a) Future signal estimate
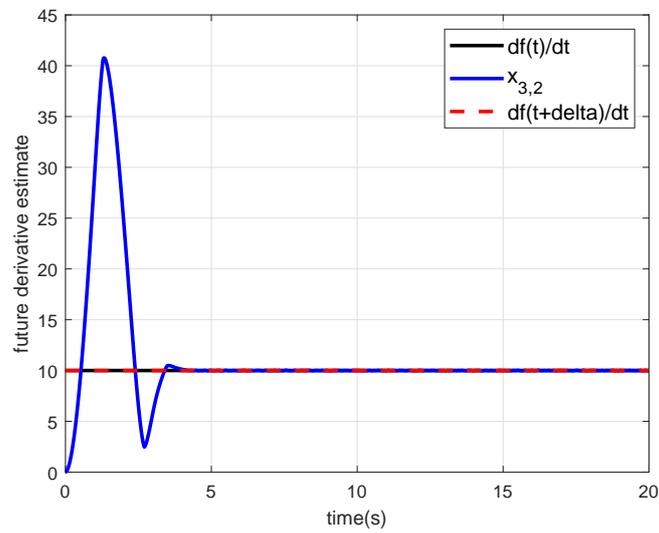

9(b) Future derivative estimate
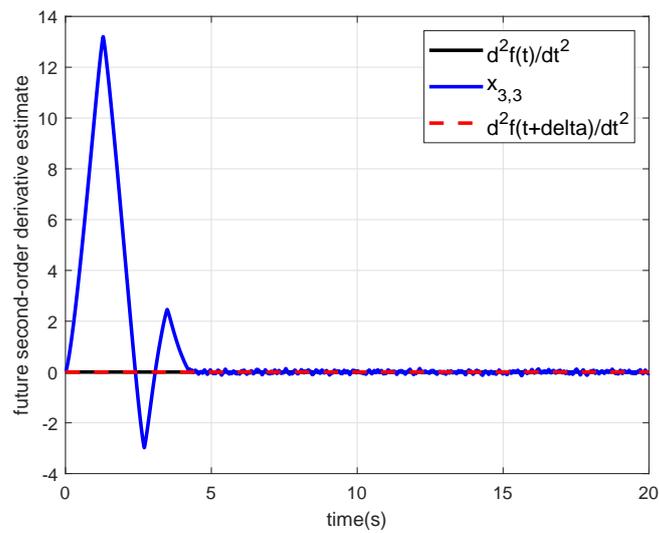

9(c) Future second-order derivative estimate

Figure 9. Nonlinear differentiator-predictor with step input signal
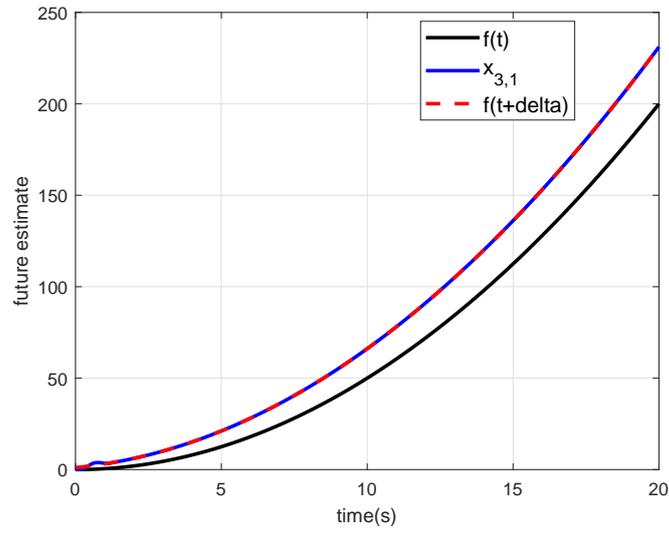
25

10(a) Future signal estimate
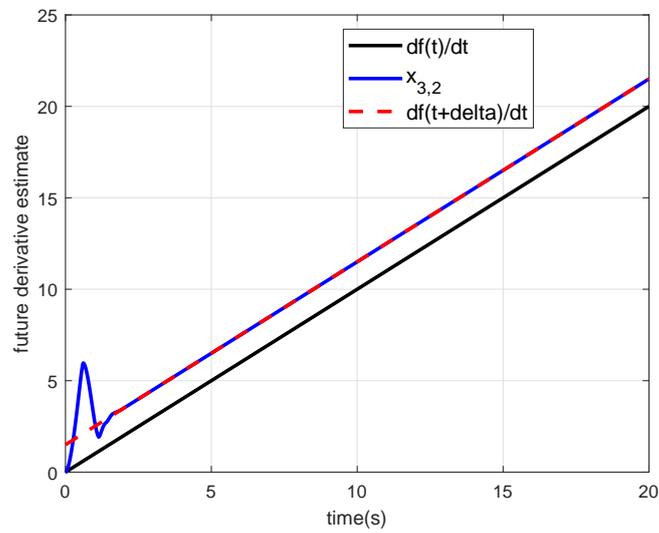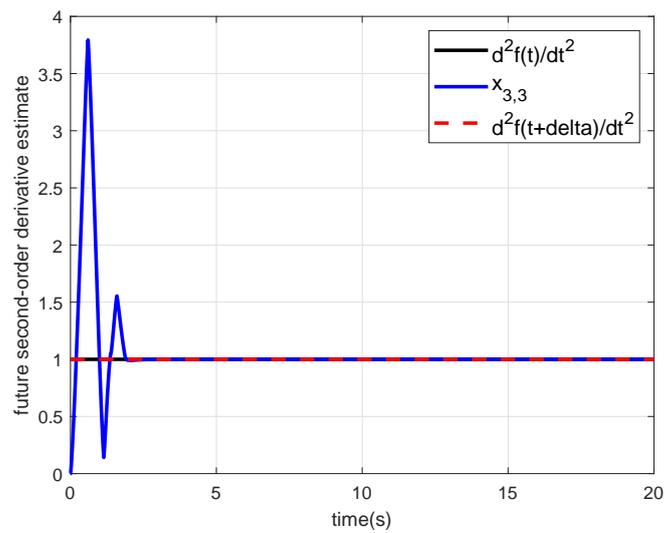

10(b) Future derivative estimate

10(c) Future second-order derivative estimate

Figure 10. Nonlinear differentiator-predictor with ramp input signal

11(a) Future signal estimate

11(b) Future derivative estimate

11(c) Future second-order derivative estimate

Figure 11. Nonlinear differentiator-predictor with parabola input signal

5.2 For slow time-varying input signal

We select the function $f(t) = \sin(0.5t) + 0.5\cos(t)$ as the input signal $f(t)$.

Suppose we can get $\max \omega \left[f(t) \right] \le \omega_f = 1.2\text{rad/s}$, and the maximal magnitude-frequency characteristic at the natural frequency is required to be $a_N = 3\text{dB}$. Therefore, $\Delta_i \le \frac{1}{\omega_f} = 0.83$ holds, and we can select $\Delta_i = 0.6s$. From the requirement $a_N = 3\text{dB}$, we can get $m \le 1.032 a_N = 3.096$. We select $m = 3$. Therefore, the future time interval $\Delta = \sum\limits_{i=1}^{3} \Delta_i = 0.6s \times 3 = 1.8s$ is assigned.

For comparison, the functions $f(t + 1.8) = \sin(0.5(t + 1.8)) + 0.5\cos(t + 1.8)$, $\dot{f}(t + 1.8) = 0.5\cos(0.5(t + 1.8)) - 0.5\sin(t + 1.8)$ and $\ddot{f}(t + 1.8) = -0.25\sin(0.5(t + 1.8)) - 0.5\cos(t + 1.8)$ are the desired future signal, the future first-order derivative and the future second-order derivative, respectively. The estimations of future signal $f(t + 1.8)$ and derivatives $f^{(j)}(t + 1.8)$ are determined from the current signal $f(t)$. Moreover, the high-frequency noise is considered in the input signal. The parameters of differentiator-predictors are shown in Table 2.

Table 2 The parameters of differentiator-predictors

differentiator-predictor	n	k_1	k_2	k_3	k_4	α_1	$1/\varepsilon$
Linear high-gain differentiator-predictor	3	6	11	6		1	$100\frac{1-e^{-0.02t}}{1+e^{-0.02t}}$ (no noise) 2.5 (consider noise)
Nonlinear differentiator-predictor	3	6	11	6		0.2	1.1
Nonlinear extended differentiator-predictor	4	1	4	6	4	0.1	1.1

*A. Simulation on linear high-gain differentiator-predictor*

Figures 12(a)-12(c) describe the estimations by the high-gain differentiator-predictor without considering of noise: $x_{3,1}$, $x_{3,2}$ and $x_{3,3}$ estimate $f(t + 1.8)$, $\dot{f}(t + 1.8)$ and $\ddot{f}(t + 1.8)$, respectively. It means that, at any time t, we can observe the values of signal and its derivative at time $t + 1.8s$. Parameter ε should be selected very small. Figures 13(a)-13(c) present the estimations of the future signal, its first-order and second-order derivatives, respectively, by considering of noise. In order to reduce noise effect, according to the frequency analysis, a relatively larger ε should be selected.

B. Simulation on nonlinear differentiator-predictor

Figures 14(a)-14(c) show the estimations by the nonlinear differentiator-predictor without considering of noise: $x_{3,1}$, $x_{3,2}$ and $x_{3,3}$ estimate $f(t + 1.8)$, $\dot{f}(t + 1.8)$ and $\ddot{f}(t + 1.8)$, respectively.The chattering still exists in the estimate outputs of future derivatives, although it has been reduced to some extent by the continuous structure: even a small chattering exists in the first step, it is amplified by the iterations. Figures 15(a)-15(c) present the future derivatives estimate by considering of noise. The parameters of the nonlinear differentiator-predictor do not need to change, and the noise effect is reduced.

*C. Simulation on nonlinear extended differentiator-predictor*

Figures 16(a)-16(c) show the estimations by the nonlinear extended differentiator-predictor: $x_{3,1}$, $x_{3,2}$ and $x_{3,3}$ estimate $f(t + 1.8)$, $\dot{f}(t + 1.8)$ and $\ddot{f}(t + 1.8)$, respectively. No chattering happens for the estimate outputs of $x_{3,1}$ and $x_{3,2}$, and the chattering in $x_{3,3}$ becomes very weak.
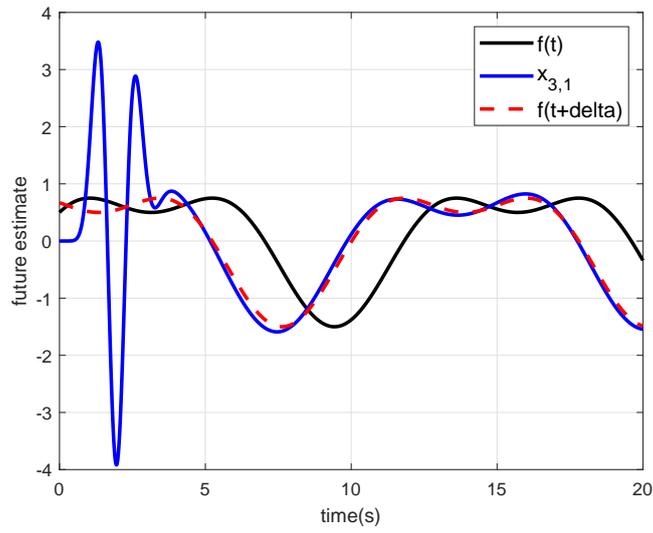
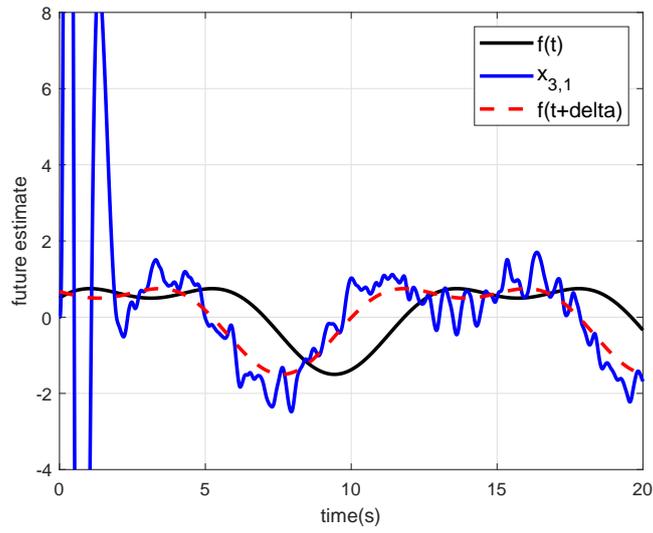12(a) Future signal estimate
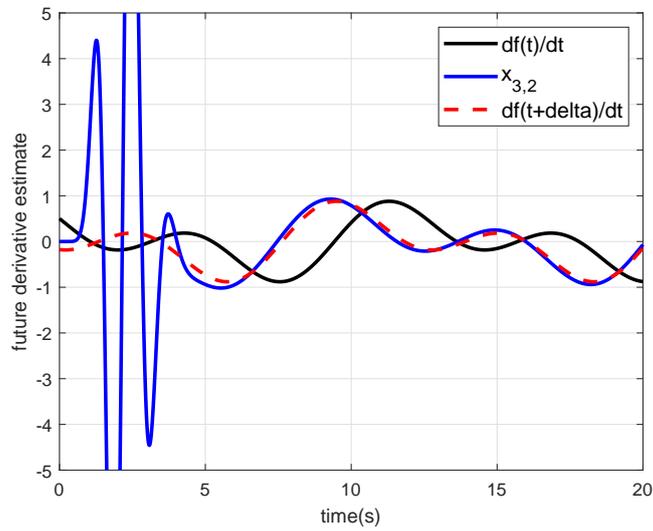
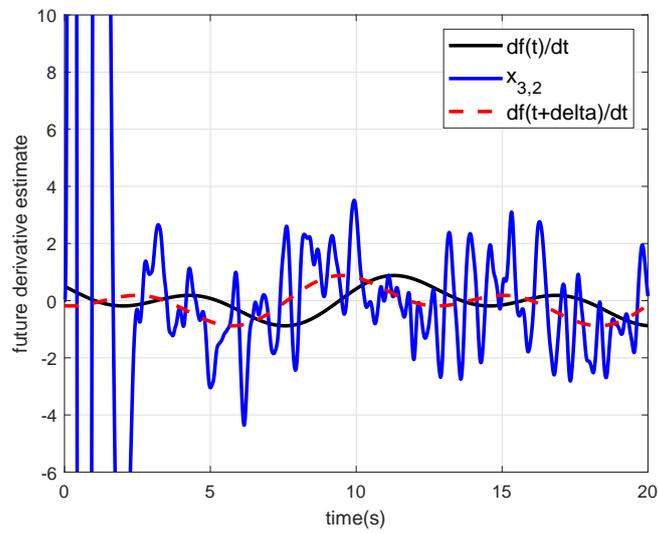
12(b) Future derivative estimate
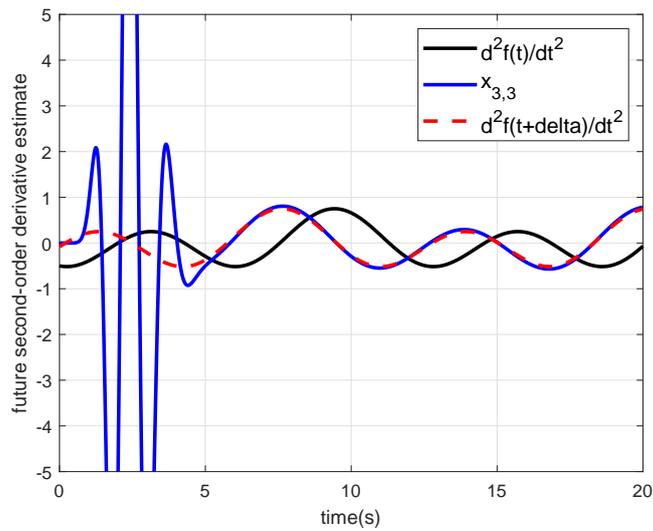
12(c) Future second-order derivative estimate
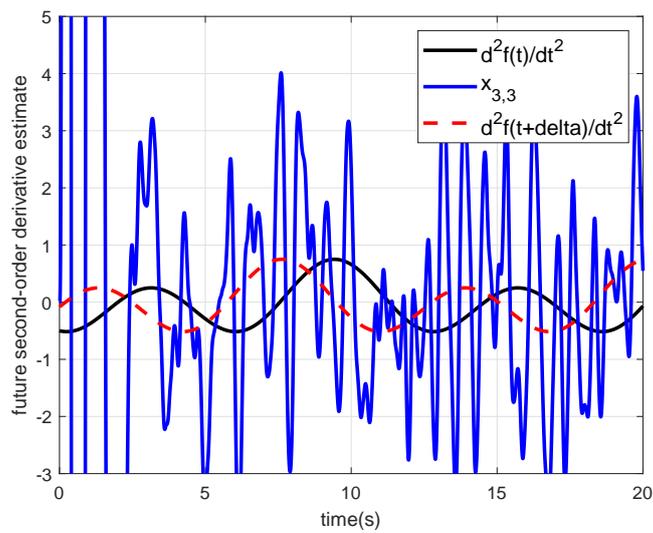
Figure 12. High-gain predictor-differentiator without noise

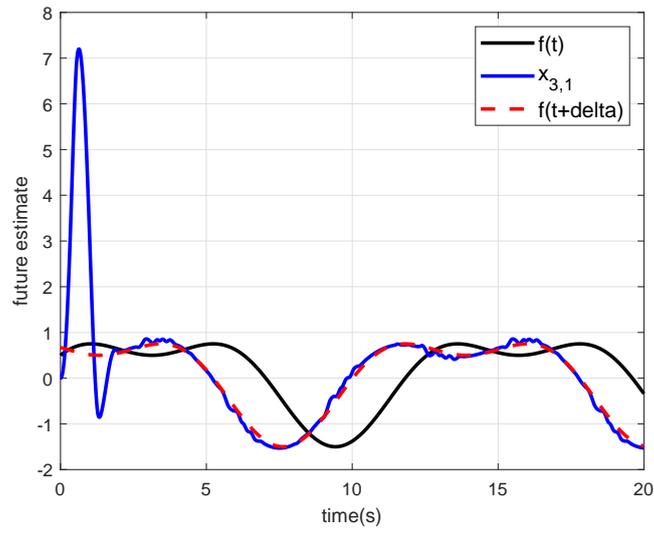13(a) Future signal estimate

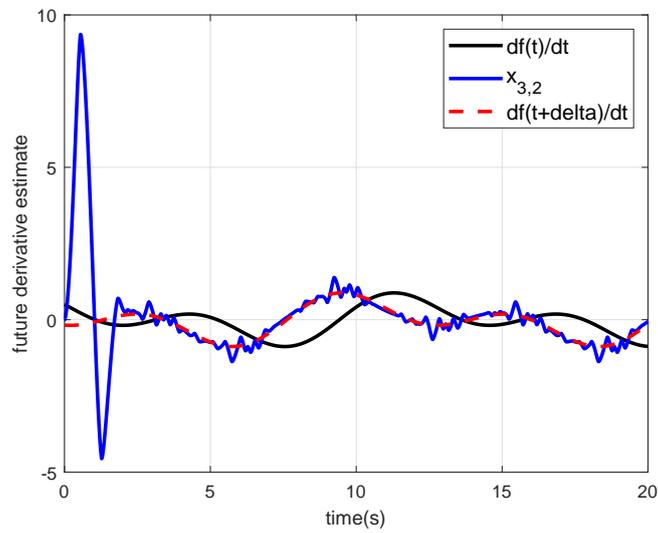13(b) Future derivative estimate

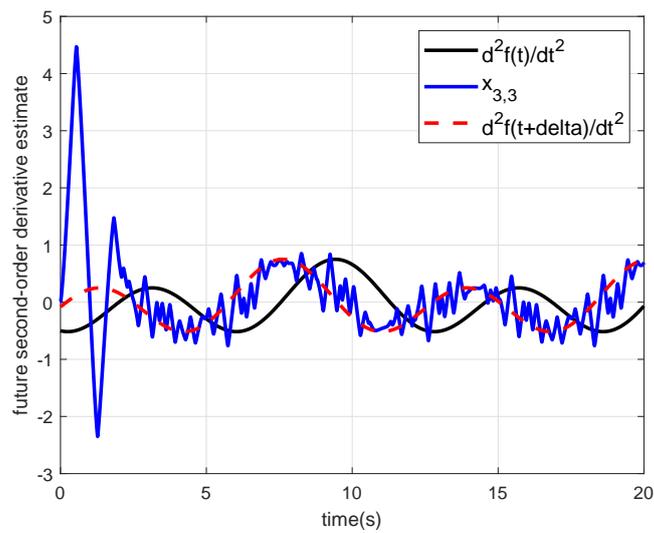13(c) Future second-order derivative estimate

Figure 13. High-gain predictor-differentiator with noise
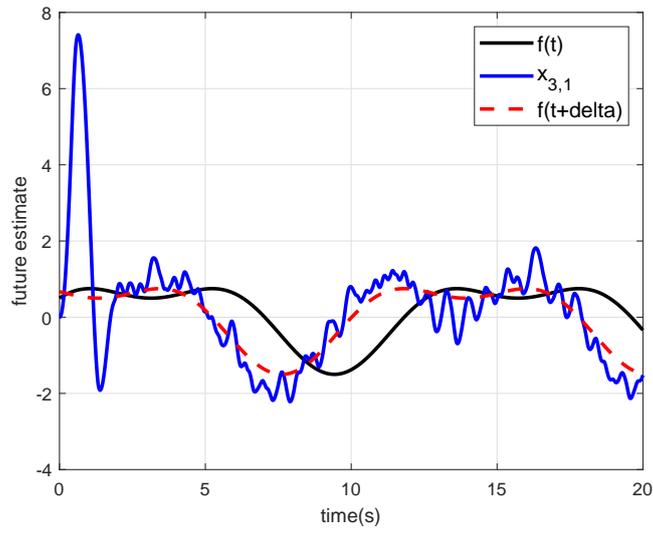
14(a) Future signal estimate
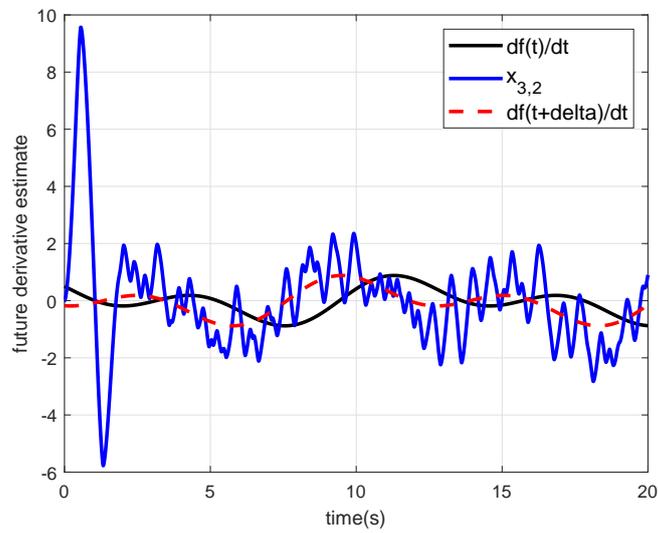

14(b) Future derivative estimate

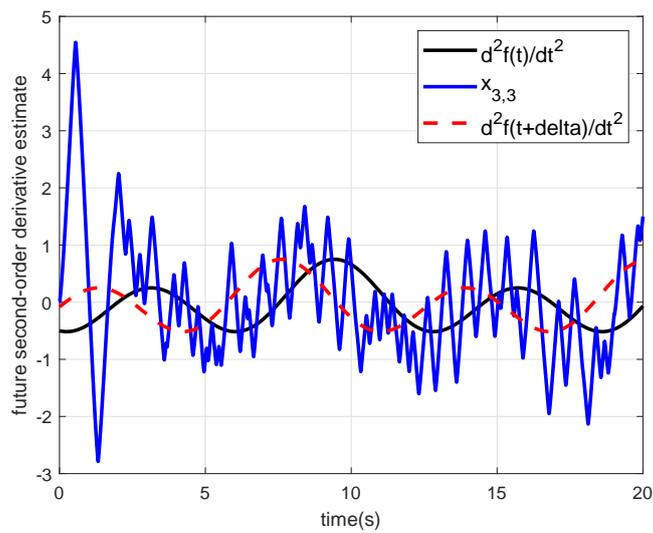14(c) Future second-order derivative estimate

Figure 14. Nonlinear predictor-differentiator without noise

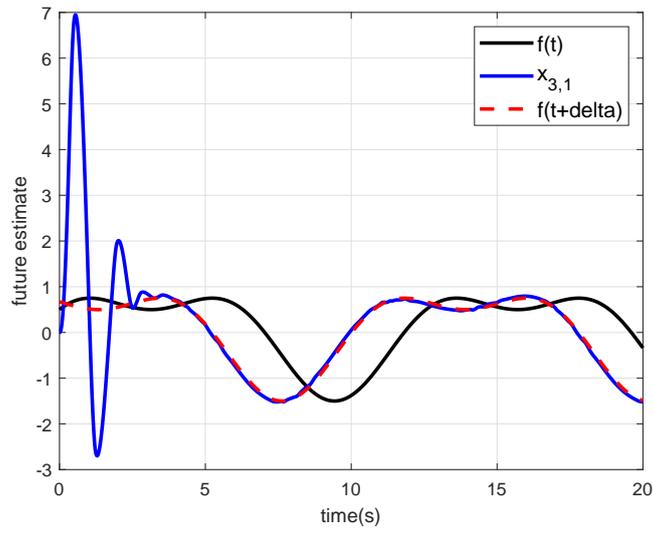15(a) Future signal estimate

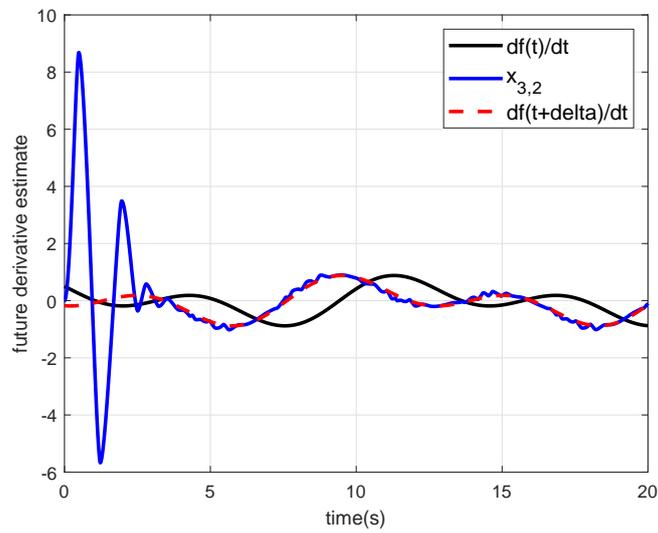15(b) Future derivative estimate
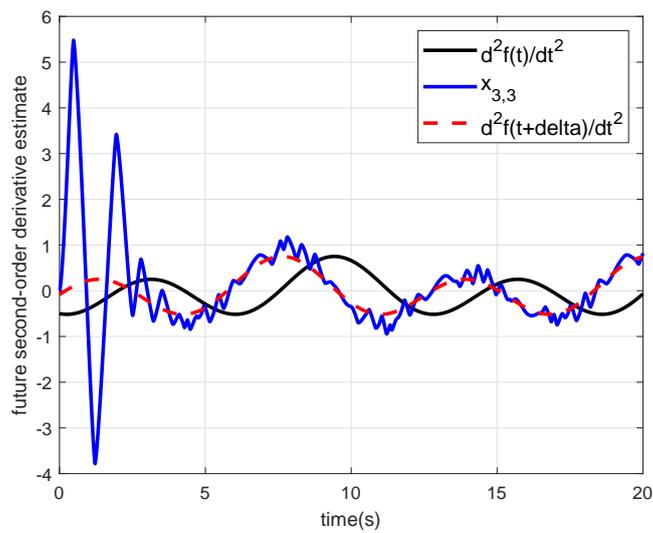

15(c) Future second-order derivative estimate

Figure 15. Nonlinear predictor-differentiator with noise

16(a) Future signal estimate

16(b) Future derivative estimate

16(c) Future second-order derivative estimate

Figure 16. Extended nonlinear predictor-differentiator

D. Performance analysis of linear and nonlinear (extended) differentiator-predictors

From the simulations, we can find that, not only the differentiator-predictors can provide the future signal, but also the future derivatives can be estimated, in spite of the requirement of large future time interval Δ. Moreover, high-frequency noise can be reduced to some extent. This confirmed the frequency analysis of the differentiator-predictors.

For the high-gain differentiator-predictor, in order to reduce noise, the parameter ε needs to increase. Otherwise, much noise will exist in the estimate outputs. Alternatively, the filters can be used before each iteration for rejecting the noise.

For the nonlinear differentiator-predictor, even noise exists, the parameter ε do not need to change. It means that the nonlinear differentiator-predictor has much stronger adaptability and robustness than linear high-gain differentiator-predictor. Also, we can find that, serious peaking phenomenon happens when the high-gain differentiator-predictor is used; while, no large peaking exists when the nonlinear differentiator-predictor is adopted. Comparing to the nonlinear differentiator-predictor, the chattering effect is reduced more sufficiently by the nonlinear extended differentiator-predictor.

E. Maximal prediction time interval

The maximal prediction time interval is decided by the input signal bandwidth and the requirement of estimate accuracy. We use the examples to explain it, and the input signal is selected as $f(t) = \sin(0.5t) + 0.5\cos(t)$.

(i) Suppose the maximal angular frequency of input signal satisfies $\max \omega [f(t)] \leq \omega_f = 1.2 \text{rad/s}$, and the maximal magnitude-frequency characteristic at the natural frequency is required to be $a = 5\text{dB}$. Therefore, $\Delta_i \leq \frac{1}{\omega_f} = 0.83$ holds, and we can select the prediction time interval at each step to be $\Delta_i = 0.6s$. From the requirement $a_N = 5\text{dB}$, we can get $\max\{m\} = int\{1.032a_N\} = 5$. Therefore, the future time interval $\Delta = \sum\limits_{i=1}^{5} \Delta_i = 0.6s \times 5 = 3s$ is assigned.

(ii) Suppose the maximal angular frequency of input signal satisfies $\max \omega [f(t)] \leq \omega_f = 1.2 \text{rad/s}$, and the maximal magnitude-frequency characteristic at the natural frequency is required to be $a_N = 6\text{dB}$. Therefore, $\Delta_i \leq \frac{1}{\omega_f} = 0.83$ holds, and we can select the prediction time interval at each step to be $\Delta_i = 0.6s$. From the requirement $a_N = 6\text{dB}$, we can get $\max\{m\} = int\{1.032a_N\} = 6$. Therefore, the future time interval $\Delta = \sum\limits_{i=1}^{6} \Delta_i = 0.6s \times 6 = 3.6s$ is assigned.

Figure17 shows the future prediction under the conditions of (i) and (ii), respectively. Due to the different requirements ($a_N = 5\text{dB}$ for case (i), and $a_N = 6\text{dB}$ for case (ii)), the estimate accuracy of case (i) is better than that of case (ii).

5.3 Fast time-varying input signal and maximal time interval prediction

We select the function $f(t) = 2\sin(0.5t) + \sin(t) + 0.5\cos(3t)$ as the input signal, in which, $\cos(3t)$ is fast time-varying element. Suppose we can know $\max \omega [f(t)] \leq \omega_f = 4\text{rad/s}$.

(i) Nonlinear differentiator-predictor (Different parameter selections). Suppose the maximal magnitude-frequency characteristic at the natural frequency is required to be $a_N = 3\text{dB}$. Therefore, $\Delta_i \leq \frac{1}{\omega_f} = 0.25$ holds, and we can select $\Delta_i = 0.2s$. From the requirement $a_N = 3\text{dB}$, we get $\max\{m\} = int\{1.032a_N\} = 3$. Therefore, the future time interval $\Delta = \sum\limits_{i=1}^{3} \Delta_i = 0.2s \times 3 = 0.6s$ is
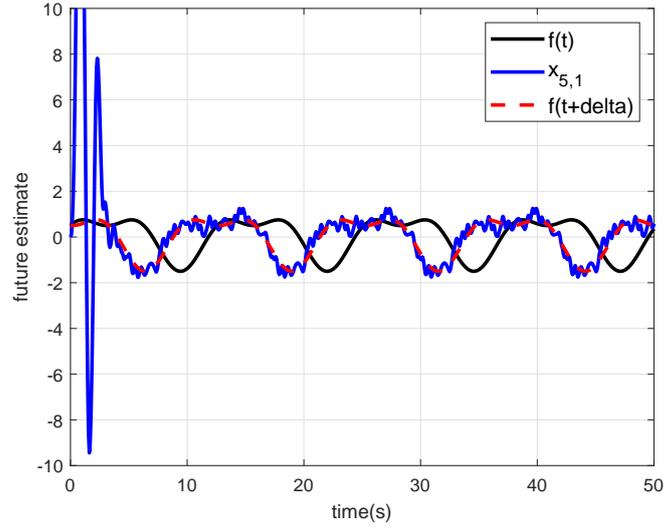
17(a) $\Delta = 3s = 0.6s \times 5$
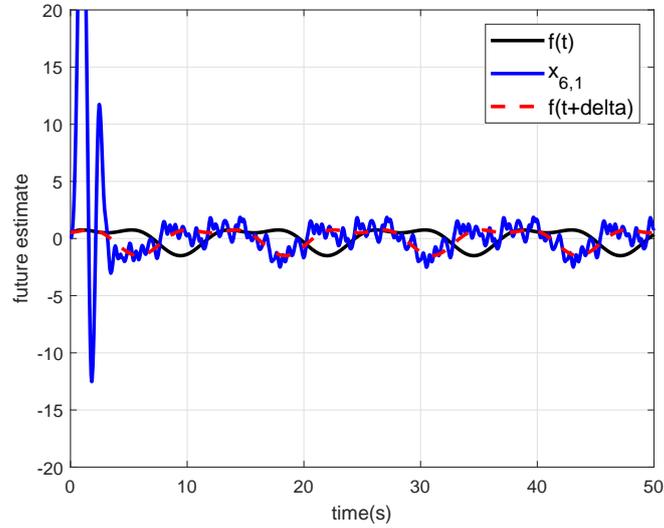
17(b) $\Delta = 3.6s = 0.6s \times 6$

Figure 17. Nonlinear differentiator-predictor with large future interval prediction

assigned. For comparison, the function $f\left(t\right) = 2\sin(0.5\left(t + 0.6\right)) + \sin(t + 0.6) + 0.5\cos(3\left(t + 0.6\right))$ is the desired future signal. Figure 18 describes the different prediction performances for the different parameter selections of ε and α_1. From Figure 18, we can find that:

Parameter $\varepsilon \in (0, 1)$ affects the system bandwidth: Decreasing ε (from 0.5 to 0.2), the low-pass bandwidth becomes larger, however, frequent oscillations start to happen. The selection of ε should satisfy that the bandwidth for the differentiator is same to that of the predictor.

Parameter $\alpha_1 \in (0, 1)$ affects the estimate accuracy: Parameter α_1 guarantees that the system is continuous; smaller α_1 can obtain more precise estimations (from 0.2 to 0.1); on the other hand, larger α_1 can reduce much noise (from 0.1 to 0.2), however, the large estimate error starts to happen.

The parameters tuning in the simulation confirms the presented rules of nonlinear differentiator-predictor parameters selection.
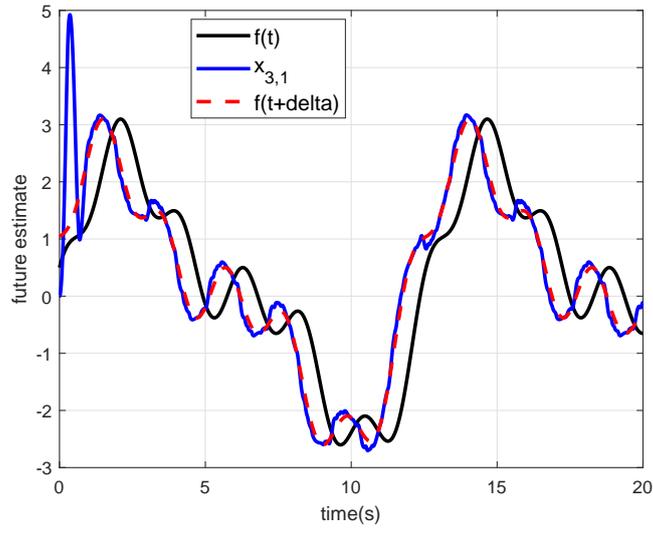
18(a) $\varepsilon = 0.5$, $\alpha_1 = 0.1$
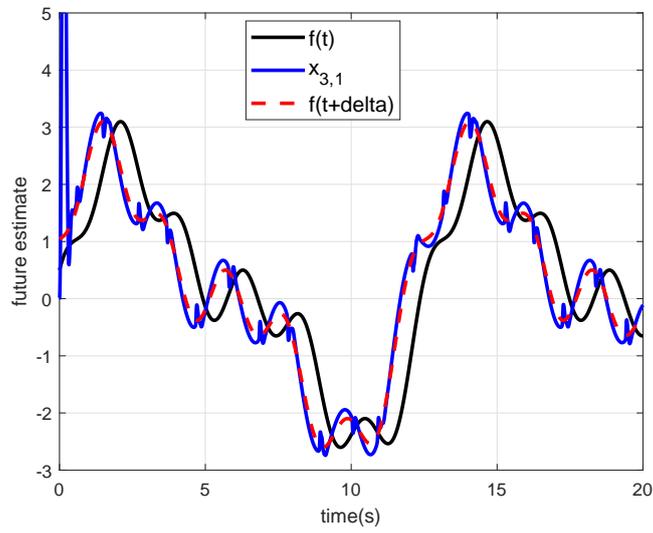
18(b) $\varepsilon = 0.2$, $\alpha_1 = 0.1$

18(c) $\varepsilon = 0.5$, $\alpha_1 = 0.2$

Figure 18. Nonlinear differentiator-predictor with fast-varying input signal: $\Delta = 0.6s = 0.2s \times 3$

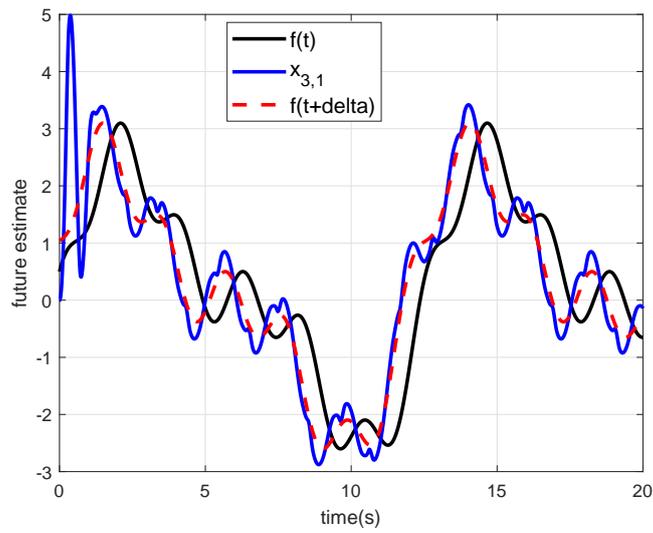(ii) Nonlinear extended differentiator-predictor. Suppose the maximal magnitude-frequency characteristic at the natural frequency is required to be $a_N = 5\text{dB}$. Therefore, $\Delta_i \leq \frac{1}{\omega_f} = 0.25$ holds, and we can select $\Delta_i = 0.2s$. From the requirement $a_N = 5\text{dB}$, we get $\max\{m\} = int\{1.032a_N\} = 5$. Therefore, the future time interval $\Delta = \sum_{i=1}^{5} \Delta_i = 0.2s \times 5 = 1s$ is assigned. Parameters $\varepsilon = 0.2$ and $\alpha_1 = 0.1$ are selected. For comparison, the function $f(t) = 2\sin(0.5(t+1)) + \sin(t+1) + 0.5\cos(3(t+1))$ is the desired future signal.

Figures 18(a) and 19 show the future prediction for the fast time-varying input signal under the conditions of (i) and (ii), respectively. The estimate accuracy decreases as the prediction time interval increases (e.g., $a_N = 3\text{dB}$ for case (i), and $a_N = 5\text{dB}$ for case (ii)).
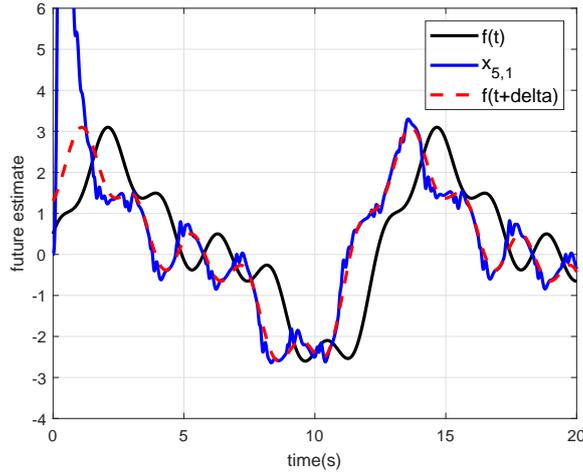
Figure 19. Nonlinear extended differentiator-predictor with fast-varying input signal:
$$\Delta = 1s = 0.2s \times 5$$

6   Conclusion

This paper presents two types of stable differentiator-predictors to estimate the future signal and its derivatives. In fact, these differentiator-predictors can also correct the delay for delayed signal, and the undelayed derivatives can be observed. The future jobs are to reduce the peaking more effectively for the high-gain differentiator-predictor, and to reject the chattering effect completely for the nonlinear differentiator-predictor.

Appendix

Proof of Theorem 1:

1) Estimation of current derivatives

For system (25), we define the errors as

$$e_{0,j} = f^{(j-1)}(t) - x_{0,j}(t)\,;\, j = 1, \cdots, n;\, e_0(t) = \left[e_{0,1}(t) \cdots e_{0,n}(t)\right]^T \tag{73}$$

Then, the estimate error system can be given by

$$\dot{e}_0(t) = \overline{A}e_0(t) + Bf^{(n)}(t) \tag{74}$$

where,

$$\overline{A} = \begin{bmatrix} -\frac{k_n}{\varepsilon} & 1 & 0 & \cdots & 0 \\ -\frac{k_{n-1}}{\varepsilon^2} & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ -\frac{k_1}{\varepsilon^n} & \cdots & \cdots & \cdots & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{n \times 1} \tag{75}$$

We know that $\overline{A}$ is a matrix in the observable standard form, and $s^n + k_n s^{n-1} + \cdots + k_2 s + k_1 = 0$ is Hurwitz. Therefore, the eigenvalues $-\lambda_j < 0$, where, $j = 1, \cdots, n$. Then, for matrix $\overline{A}$, there exist an invertible matrix T, such that

$$\overline{A} = T \cdot \mathrm{diag}\left\{ -\frac{\lambda_1}{\varepsilon}, \cdots, -\frac{\lambda_n}{\varepsilon} \right\} T^{-1} \tag{76}$$

From the properties of matrix $\overline{A}$, we know that $\|T\|$ and $\|T^{-1}\|$ are bounded without involving ε. The solution to the system (74) is

$$e_0(t) = e^{\overline{A}t} e_0(0) + \int_0^t e^{\overline{A}(t-\tau)} B f^{(n)}(\tau) \, d\tau \tag{77}$$

Taking the norm for (77), we get

$$\|e_0(t)\| \le \left\| e^{\overline{A}t} \right\| \|e_0(0)\| + \int_0^t \left\| e^{\overline{A}(t-\tau)} \right\| \|B\| \left| f^{(n)}(\tau) \right| d\tau \tag{78}$$

According to Taylor expansion and the properties of matrix $\overline{A}$, we know that

$$e^{\overline{A}t} = \sum_{i=0}^{\infty} \frac{1}{i!} \overline{A}^i t^i = T \left(\sum_{i=0}^{\infty} \frac{1}{i!} \left(\mathrm{diag}\left\{ -\frac{\lambda_1}{\varepsilon}, \cdots, -\frac{\lambda_n}{\varepsilon} \right\} \right)^i t^i \right) T^{-1}$$
$$= T \cdot \mathrm{diag}\left\{ e^{-\frac{\lambda_1}{\varepsilon}t}, \cdots, e^{-\frac{\lambda_n}{\varepsilon}t} \right\} T^{-1} \tag{79}$$

Define $\lambda = \min_{j \in \{1, \cdots, n\}} \{\lambda_j\}$, $\|e_0(0)\| \le h_0$, and $\left| f^{(n)}(\tau) \right| \le h_n$, where, h_0 and h_n and are the bounded constants. Then, according to (79), (78) can be written by

$$\|e_0(t)\| \le \left\| T \cdot \mathrm{diag}\left\{ e^{-\frac{\lambda_1}{\varepsilon}t}, \cdots, e^{-\frac{\lambda_n}{\varepsilon}t} \right\} T^{-1} \right\| \|e_0(0)\|$$
$$+ \int_0^t \left\| T \cdot \mathrm{diag}\left\{ e^{-\frac{\lambda_1}{\varepsilon}(t-\tau)}, \cdots, e^{-\frac{\lambda_n}{\varepsilon}(t-\tau)} \right\} T^{-1} \right\| \|B\| \left| f^{(n)}(\tau) \right| d\tau$$
$$\le \|T\| \left\| T^{-1} \right\| \left(h_0 e^{-\frac{\lambda}{\varepsilon}t} + h_n \int_0^t e^{-\frac{\lambda}{\varepsilon}(t-\tau)} d\tau \right)$$
$$= \|T\| \left\| T^{-1} \right\| \left(h_0 e^{-\frac{\lambda}{\varepsilon}t} + h_n \frac{\varepsilon}{\lambda} \left(1 - e^{-\frac{\lambda}{\varepsilon}t} \right) \right) \tag{80}$$

Therefore, taking limit about time t for (80), we can get

$$\lim_{t\to\infty} \|e_0(t)\| \le \varepsilon \frac{h_n}{\lambda} \|T\| \|T^{-1}\| \tag{81}$$

Define $l_0 = \frac{h_n}{\lambda} \|T\| \|T^{-1}\|$, then, (81) is rewritten by

$$\lim_{t\to\infty} \|e_0(t)\| \le \varepsilon l_0 \tag{82}$$

i.e.,

$$\lim_{t\to\infty} \left| f^{(j-1)}(t) - x_{0,j}(t) \right| \le \varepsilon l_0; \quad j = 1, \cdots, n \tag{83}$$

2) Iterative estimate of future signal and future derivatives

According to $A^n = O_{n\times n}$, the Taylor expansion of $e^{A\Delta_i}$ can be given by

$$e^{A\Delta_i} = \sum_{j=1}^{n} \frac{1}{(j-1)!} (A\Delta_i)^{j-1}$$

$$= \begin{bmatrix} 1 & \Delta_i & \frac{1}{2!}\Delta_i^2 & \cdots & \frac{1}{(n-1)!}\Delta_i^{n-1} \\ 0 & 1 & \Delta_i & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & \frac{1}{2!}\Delta_i^2 \\ \vdots & \vdots & \ddots & \ddots & \Delta_i \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{84}$$

Therefore, the following relation holds:

$$Ce^{A\Delta_i} X_{i-1}(t) = \sum_{j=1}^{n} \frac{1}{(j-1)!} \Delta_i^{j-1} x_{i-1,j}(t) \tag{85}$$

Define $x_{i-1,j}(t) \overset{\text{def}}{=} x_{i-1,1}^{(j-1)}(t)$ (where, $j = n+1, \cdots$) to compensate the dimension of $X_{i-1}(t)$ into infinity. According to Taylor expansion and the element derivative relations in $X_{i-1}(t)$ and its expansion, we know that

$$x_{i-1,1}(t + \Delta_i) = \sum_{j=1}^{\infty} \frac{1}{(j-1)!} \Delta_i^{j-1} x_{i-1,j}(t) \tag{86}$$

Furthermore, $x_{i-1,1}(t)$ in $X_{i-1}(t)$ can be expressed by the Fourier form, as follows:

$$x_{i-1,1}(t) = a_0 + \sum_{k=1}^{\infty} a_k \sin(k\omega_0 t) + \sum_{k=1}^{\infty} b_k \cos(k\omega_0 t) \tag{87}$$

where, ω_0 is the base frequency in $x_{i-1,1}(t)$. Therefore, the n-th derivative of $x_{i-1,1}(t)$ is

$$x_{i-1,n+1}(t) = x_{i-1,1}^{(n)}(t) = \sum_{k=1}^{\infty} \bar{a}_k (k\omega_0)^n \sin(k\omega_0 t) + \sum_{k=1}^{\infty} \bar{b}_k (k\omega_0)^n \cos(k\omega_0 t) \tag{88}$$

Therefore, we get

$$\frac{\Delta_i^n}{n!} x_{i-1,n+1}(t) = \sum_{k=1}^{\infty} \bar{a}_k \frac{(\Delta_i k \omega_0)^n}{n!} \sin(k\omega_0 t) + \sum_{k=1}^{\infty} \bar{b}_k \frac{(\Delta_i k \omega_0)^n}{n!} \cos(k\omega_0 t) \tag{89}$$

Define:

$$O\left(\frac{(\Delta_i k \omega_0)^n}{n!}\right) = \sum_{j=n+1}^{\infty} \frac{1}{(j-1)!} \Delta_i^{j-1} x_{i-1,j}(t) \tag{90}$$

and we define $\left| O\left(\frac{(\Delta_i k \omega_0)^n}{n!}\right) \right| \le L_{\Delta i}$. Therefore, from (85), (86) and (90), we get

$$x_{i-1,1}(t + \Delta_i) = C e^{A\Delta_i} X_{i-1}(t) + O\left(\frac{(\Delta_i k \omega_0)^n}{n!}\right) \tag{91}$$

We rewrite (26) by

$$\dot{X}_i(t) = A X_i(t) + K(\varepsilon)\left(C e^{A\Delta_i} X_{i-1}(t) - C X_i(t)\right) \tag{92}$$

where, $i = 1, \cdots, m$. From (28), the following relation holds:

$$C X_i(t) = x_{i,1}(t) \tag{93}$$

Then, considering of (91) and (93), system (92) can be written by

$$\dot{X}_i(t) = A X_i(t) + K(\varepsilon)\left(x_{i-1,1}(t + \Delta_i) - O\left(\frac{(\Delta_i k \omega_0)^n}{n!}\right) - x_{i,1}(t)\right) \tag{94}$$

where, $i = 1, \cdots, m$. Therefore, according to the result (18) in Lemma 3, for system (94), there exist the bounded constants l_i and d_i, such that

$$\lim_{t \to \infty} \left\| \begin{bmatrix} x_{i-1,1}(t+\Delta_i) \\ \dot{x}_{i-1,1}(t+\Delta_i) \\ \vdots \\ x_{i-1,1}^{(n-1)}(t+\Delta_i) \end{bmatrix} - \begin{bmatrix} x_{i,1}(t) \\ x_{i,2}(t) \\ \vdots \\ x_{i,n}(t) \end{bmatrix} \right\| \le \varepsilon l_i + \frac{L_{\Delta i} d_i}{\varepsilon^{n-1}}; \ i = 1, \cdots, m \tag{95}$$

Then, according to the variable relations in (26), we get

$$\lim_{t \to \infty} \| X_{i-1}(t + \Delta_i) - X_i(t) \| \le \varepsilon l_i + \frac{L_{\Delta i} d_i}{\varepsilon^{n-1}}; \ i = 1, \cdots, m \tag{96}$$

Thus, for each variable in the estimate vectors, it follows that

$$\lim_{t \to \infty} |(x_{i-1,j}(t + \Delta_i)) - x_{i,j}(t)| \le \varepsilon l_i + \frac{L_{\Delta i} d_i}{\varepsilon^{n-1}}; \ j = 1, \cdots, n; \ i = 1, \cdots, m \tag{97}$$

3) Combination of 1) and 2)

Using $t + \Delta_1$ instead of t, the result (83) for system (25) still holds, and it can be written by

$$\lim_{t \to \infty} \left| f^{(j-1)}(t + \Delta_1) - x_{0,j}(t + \Delta_1) \right| \le \varepsilon l_0 \tag{98}$$

40

Then, from (97) and (98), and considering of $-x_{0,j}(t+\Delta_1) + x_{0,j}(t+\Delta_1) = 0$, we can get

$$\lim_{t\to\infty} \left| f^{(j-1)}(t+\Delta_1) - x_{1,j}(t) \right|$$

$$= \lim_{t\to\infty} \left| f^{(j-1)}(t+\Delta_1) - x_{0,j}(t+\Delta_1) + x_{0,j}(t+\Delta_1) - x_{1,j}(t) \right|$$

$$\leq \lim_{t\to\infty} \left| f^{(j-1)}(t+\Delta_1) - x_{0,j}(t+\Delta_1) \right| + \lim_{t\to\infty} \left| x_{0,j}(t+\Delta_1) - x_{1,j}(t) \right|$$

$$\leq \varepsilon l_0 + \varepsilon l_1 + \frac{L_{\Delta 1} d_1}{\varepsilon^{n-1}} = \varepsilon(l_0 + l_1) + \frac{L_{\Delta 1} d_1}{\varepsilon^{n-1}} \tag{99}$$

Furthermore, considering of $-x_{1,j}(t+\Delta_2) + x_{1,j}(t+\Delta_2) = 0$, it follows that

$$\lim_{t\to\infty} \left| f^{(j-1)}(t+\Delta_1+\Delta_2) - x_{2,j}(t) \right|$$

$$\leq \lim_{t\to\infty} \left| f^{(j-1)}(t+\Delta_2+\Delta_1) - x_{1,j}(t+\Delta_2) \right| + \lim_{t\to\infty} \left| x_{1,j}(t+\Delta_2) - x_{2,j}(t) \right|$$

$$\leq \varepsilon(l_0+l_1) + \frac{L_{\Delta 1} d_1}{\varepsilon^{n-1}} + \varepsilon l_2 + \frac{L_{\Delta 2} d_2}{\varepsilon^{n-1}} = \varepsilon(l_0+l_1+l_2) + \frac{1}{\varepsilon^{n-1}}(L_{\Delta 1} d_1 + L_{\Delta 2} d_2) \tag{100}$$

In general, iterating i from 1 to m, and considering of $-x_{m-1,j}(t+\Delta_m) + x_{m-1,j}(t+\Delta_m) = 0$, we get

$$\lim_{t\to\infty} \left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - x_{m,j}(t) \right|$$

$$= \lim_{t\to\infty} \left| f^{(j-1)}\left(t+\sum_{i=1}^{m-1}\Delta_i+\Delta_m\right) - x_{m-1,j}(t+\Delta_m) + x_{m-1,j}(t+\Delta_m) - x_{m,j}(t) \right|$$

$$\leq \lim_{t\to\infty} \left| f^{(j-1)}\left(t+\Delta_m+\sum_{i=1}^{m-1}\Delta_i\right) - x_{m-1,j}(t+\Delta_m) \right| + \lim_{t\to\infty} \left| x_{m-1,j}(t+\Delta_m) - x_{m,j}(t) \right|$$

$$\leq \varepsilon \sum_{i=0}^{m-1} l_i + \frac{1}{\varepsilon^{n-1}} \sum_{k=1}^{m-1} L_{\Delta k} d_k + \varepsilon l_m + \frac{L_{\Delta m}}{\varepsilon^{n-1}} d_m = \varepsilon \sum_{i=0}^{m} l_i + \frac{1}{\varepsilon^{n-1}} \sum_{k=1}^{m} L_{\Delta k} d_k \leq \varepsilon c_1 + c_2 \frac{L_d}{\varepsilon^{n-1}} \tag{101}$$

where, $c_1 = \sum_{i=0}^{m} l_i$, $c_2 = \sum_{k=1}^{m} d_k$, $L_d = \max_{k\in\{1,\cdots,m\}}\{L_{\Delta k}\}$, and $j = 1, \cdots, n$.

Alternatively, by substituting $-\sum_{k=1}^{m} x_{k-1,j}\left(t+\sum_{i=k}^{m}\Delta_i\right) + \sum_{k=1}^{m} x_{k-1,j}\left(t+\sum_{i=k}^{m}\Delta_i\right) = 0$ into

$\lim_{t\to\infty} \left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - x_{m,j}(t) \right|$, we can also get the estimate error up-boundness (101) by using the following method:

$$\lim_{t\to\infty} \left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - x_{m,j}(t) \right|$$

$$= \lim_{t\to\infty} \left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - \sum_{k=1}^{m} x_{k-1,j}\left(t+\sum_{i=k}^{m}\Delta_i\right) + \sum_{k=1}^{m} x_{k-1,j}\left(t+\sum_{i=k}^{m}\Delta_i\right) - x_{m,j}(t) \right|$$

$$= \lim_{t \to \infty} \left| f^{(j-1)} \left(t + \sum_{i=1}^{m} \Delta_i \right) - x_{0,j} \left(t + \sum_{i=1}^{m} \Delta_i \right) \right.$$

$$\left. - \sum_{k=2}^{m} x_{k-1,j} \left(t + \sum_{i=k}^{m} \Delta_i \right) + \sum_{k=1}^{m-1} x_{k-1,j} \left(t + \sum_{i=k}^{m} \Delta_i \right) + x_{m-1,j} \left(t + \Delta_m \right) - x_{m,j} \left(t \right) \right|$$

$$= \lim_{t \to \infty} \left| f^{(j-1)} \left(t + \sum_{i=1}^{m} \Delta_i \right) - x_{0,j} \left(t + \sum_{i=1}^{m} \Delta_i \right) \right.$$

$$\left. + \sum_{k=1}^{m-1} \left\{ x_{k-1,j} \left(t + \sum_{i=k}^{m} \Delta_i \right) - x_{k,j} \left(t + \sum_{i=k+1}^{m} \Delta_i \right) \right\} + x_{m-1,j} \left(t + \Delta_m \right) - x_{m,j} \left(t \right) \right|$$

$$\leq \lim_{t \to \infty} \left| f^{(j-1)} \left(t + \sum_{i=1}^{m} \Delta_i \right) - x_{0,j} \left(t + \sum_{i=1}^{m} \Delta_i \right) \right|$$

$$+ \sum_{k=1}^{m-1} \lim_{t \to \infty} \left| x_{k-1,j} \left(t + \sum_{i=k+1}^{m} \Delta_i + \Delta_k \right) - x_{k,j} \left(t + \sum_{i=k+1}^{m} \Delta_i \right) \right|$$

$$+ \lim_{t \to \infty} \left| x_{m-1,j} \left(t + \Delta_m \right) - x_{m,j} \left(t \right) \right|$$

$$\leq \varepsilon l_0 + \varepsilon \sum_{k=1}^{m-1} l_i + \frac{1}{\varepsilon^{n-1}} \sum_{k=1}^{m-1} L_{\Delta k} d_k + \varepsilon l_m + \frac{L_{\Delta m}}{\varepsilon^{n-1}} d_m$$

$$= \varepsilon \sum_{i=0}^{m} l_i + \frac{1}{\varepsilon^{n-1}} \sum_{k=1}^{m} L_{\Delta k} d_k \leq \varepsilon c_1 + c_2 \frac{L_d}{\varepsilon^{n-1}} \tag{102}$$

where, $j = 1, \cdots, n$.

According to Lemma 3:

1) function $\varepsilon c_1 + c_2 \frac{L_d}{\varepsilon^{n-1}}$ in (102) has a global minimum at $\varepsilon = \left[(n-1) c_2 L_d / c_1 \right]^{(1/n)} \overset{\text{def}}{=} c_a L_d^{1/n}$, and

$$\min_{\varepsilon > 0} \left\{ \varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2 \right\} = \left(c_1 c_a + c_2 / c_a^{(r-1)} \right) L_d^{1/n} \overset{\text{def}}{=} k_a L_d^{1/n} \tag{103}$$

2) For $\varepsilon \in \left[c_a L_d^{1/n}, 1 \right)$, $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2$ is a strictly increasing function of ε, and $\varepsilon c_1 + \frac{L_d}{\varepsilon^{n-1}} c_2 \leq L \varepsilon$, where $L = c_1 + c_2 / c_a^n$. This concludes the proof. ∎

Proof of Theorem 2:

1) Estimation of current derivatives

For (32), according to Lemma 4, there exist $\gamma > 1$, time $t_0 > 0$ and a bounded constant $l_0 > 0$, such that, for $t \geq t_0$,

$$\left| f^{(j-1)} \left(t \right) - x_{0,j} \left(t \right) \right| \leq l_0 \varepsilon^{n\gamma - j + 1}; \ j = 1, \ldots, n \tag{104}$$

2) Iterative estimate of future signal and future derivatives

From (36), we can get

$$C e^{A \Delta_i} X_{i-1} \left(t \right) = \sum_{j=1}^{n} \frac{1}{(j-1)!} \Delta_i^{j-1} x_{i-1,j} \left(t \right); C X_i \left(t \right) = x_{i,1} \left(t \right) \tag{105}$$

Define $x_{i-1,j}(t) \stackrel{\text{def}}{=} x_{i-1,1}^{(j-1)}(t)$ (where, $j = n+1, \cdots$) to compensate the dimension of $X_{i-1}(t)$ into infinity. Meanwhile, according to Taylor expansion and the element derivative relations in $X_{i-1}(t)$ and its expansion, we know that

$$x_{i-1,1}(t + \Delta_i) = \sum_{j=1}^{\infty} \frac{1}{(j-1)!} \Delta_i^{j-1} x_{i-1,j}(t) \tag{106}$$

Therefore, from (105) and (106), the following relation holds:

$$x_{i-1,1}(t + \Delta_i) = Ce^{A\Delta_i} X_{i-1}(t) + O\left(\frac{(\Delta_i k\omega_0)^n}{n!}\right) \tag{107}$$

where, $O\left(\frac{(\Delta_i k\omega_0)^n}{n!}\right) = \sum_{j=n+1}^{\infty} \frac{1}{(j-1)!} \Delta_i^{j-1} x_{i-1,j}(t)$. We define $\left|O\left(\frac{\Delta_i^n}{n!}\right)\right| \leq L_{\Delta i}$. The system (33) can be written by

$$\dot{x}_{i,j}(t) = x_{i,j+1}(t) + \frac{k_{n-j+1}}{\varepsilon^j} \left|x_{i-1,1}(t + \Delta_i) - O\left(\frac{(\Delta_i k\omega_0)^n}{n!}\right) - x_{i,1}(t)\right|^{\alpha_{n-j+1}}$$

$$\times \text{sign}\left(x_{i-1,1}(t + \Delta_i) - O\left(\frac{(\Delta_i k\omega_0)^n}{n!}\right) - x_{i,1}(t)\right);$$

$$j = 1, \cdots, n-1$$

$$\dot{x}_{i,n}(t) = \frac{k_1}{\varepsilon^n} \left|x_{i-1,1}(t + \Delta_i) - O\left(\frac{(\Delta_i k\omega_0)^n}{n!}\right) - x_{i,1}(t)\right|^{\alpha_1}$$

$$\times \text{sign}\left(x_{i-1,1}(t + \Delta_i) - O\left(\frac{(\Delta_i k\omega_0)^n}{n!}\right) - x_{i,1}(t)\right);$$

$$i = 1, \cdots, m \tag{108}$$

For (108), from Lemma 4, there exist $\gamma > 1$, time $t_i > 0$ and bounded constants $l_i > 0$, such that, for $t \geq t_i$,

$$\left|x_{i-1,1}^{(j-1)}(t + \Delta_i) - x_{i,j}(t)\right| \leq l_i(\delta_{i,j})^{\gamma} \tag{109}$$

where, $\delta_{i,j} = \varepsilon^{n-\frac{j-1}{\gamma}} + cL_{\Delta i}^{\alpha_p} \varepsilon^{-\frac{j-1}{\gamma}}$, $j = 1, \ldots, n$, and $i = 1, \ldots, m$. From the system variable relations in (33), we know that $x_{i-1,1}^{(j-1)}(t + \Delta_i) = x_{i-1,j}(t + \Delta_i)$, $j = 1, \ldots, n$; $i = 1, \ldots, m$. Therefore, (109) can be written by

$$|x_{i-1,j}(t + \Delta_i) - x_{i,j}(t)| \leq l_i(\delta_{i,j})^{\gamma} \tag{110}$$

where, $j = 1, \ldots, n$, and $i = 1, \ldots, m$.

3) Combination of 1) and 2)

By substituting $-\sum_{k=1}^{m} x_{k-1,j}\left(t + \sum_{i=k}^{m} \Delta_i\right) + \sum_{k=1}^{m} x_{k-1,j}\left(t + \sum_{i=k}^{m} \Delta_i\right) = 0$ into

$\left|f^{(j-1)}\left(t + \sum_{i=1}^{m} \Delta_i\right) - x_{m,j}(t)\right|$, and from Lemma 4, we can get the estimate error up-boundness

43

as follows:

$$\left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - x_{m,j}(t)\right|$$

$$= \left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - \sum_{k=1}^{m} x_{k-1,j}\left(t+\sum_{i=k}^{m}\Delta_i\right) + \sum_{k=1}^{m} x_{k-1,j}\left(t+\sum_{i=k}^{m}\Delta_i\right) - x_{m,j}(t)\right|$$

$$\leq \left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - x_{0,j}\left(t+\sum_{i=1}^{m}\Delta_i\right)\right|$$

$$+ \sum_{k=1}^{m-1}\left| x_{k-1,j}\left(t+\sum_{i=k+1}^{m}\Delta_i+\Delta_k\right) - x_{k,j}\left(t+\sum_{i=k+1}^{m}\Delta_i\right)\right|$$

$$+ \left|x_{m-1,j}(t+\Delta_m) - x_{m,j}(t)\right|$$

$$\leq l_0\varepsilon^{n\gamma-j+1} + \sum_{k=1}^{m-1} l_k(\delta_{k,j})^{\gamma} + l_m(\delta_{m,j})^{\gamma} = l_0\varepsilon^{n\gamma-j+1} + \sum_{i=1}^{m} l_i(\delta_{i,j})^{\gamma}$$

$$\leq l_0\varepsilon^{n\gamma-j+1} + c_2(\delta_{dj})^{\gamma} \tag{111}$$

where, $\delta_{dj} = \max_{i\in\{1,\cdots,m\}}\{\delta_{i,j}\} = \varepsilon^{n-\frac{j-1}{\gamma}} + cL_d^{\alpha_p}\varepsilon^{-\frac{j-1}{\gamma}}$, $L_d = \max_{k\in\{1,\cdots,m\}}\{L_{\Delta k}\}$, $c_2 = \sum_{i=1}^{m} l_i$, and $j = 1,\cdots,n$.

According to Lemma 5, function $\delta_{dj} = \varepsilon^{n-\frac{j-1}{\gamma}} + cL_d^{\alpha_p}\varepsilon^{-\frac{j-1}{\gamma}}$ in (111) has the following properties:

1) δ_{dj} has a global minimum at $\varepsilon = \left(c\frac{j-1}{n\gamma-j+1}L_d^{\alpha_p}\right)^{(1/n)} \overset{\text{def}}{=} c_a L_d^{\alpha_p/n}$, and

$$\min_{\varepsilon>0}\{\delta_{dj}\} = \left(c_a^{n-\frac{j-1}{\gamma}} + c/c_a^{\frac{j-1}{\gamma}}\right)\left(L_d^{\alpha_p}\right)^{\left(1-\frac{j-1}{n\gamma}\right)} \overset{\text{def}}{=} k_a\left(L_d^{\alpha_p}\right)^{\left(1-\frac{j-1}{n\gamma}\right)}.$$

2) For $\varepsilon \geq c_a L_d^{\alpha_p/n}$, δ_{dj} is a strictly increasing function of ε, and $\delta_{dj} \leq k_b\varepsilon^{n-\frac{j-1}{\gamma}}$, where $k_b = (1+1/c_a^n)$. Therefore, $\varepsilon \in (\varepsilon_*, 1)$ is selected, where $\varepsilon_* = c_a L_d^{\alpha_p/n}$. Then, for (111), we get

$$\left| f^{(j-1)}\left(t+\sum_{i=1}^{m}\Delta_i\right) - x_{m,j}(t)\right| \leq l_0\varepsilon^{n\gamma-j+1} + c_2(\delta_{dj})^{\gamma} \leq l_0\varepsilon^{n\gamma-j+1} + c_2\left(k_b\varepsilon^{n-\frac{j-1}{\gamma}}\right)^{\gamma}$$

$$= \left(l_0 + c_2 k_b^{\gamma}\right)\varepsilon^{n\gamma-j+1} \overset{\text{def}}{=} L\varepsilon^{n\gamma-j+1} \tag{112}$$

This concludes the proof. ∎

References

Ahmed-Ali, T., Giri, F., Krstic, M., & Kahelras, M. (2018). PDE based observer design for nonlinear systems with large output delay, Syst. Control Lett., vol. 113, 1-8.

Ahrens, J. H., & Khalil, H. K. (2009). High-gain observers in the presence of measurement noise: A switched-gain approach, Automatica, vol. 45, 936-943.

Akcal, M. U., & Ure, N. K. (2017). Predictive Missile Guidance with Online Trajectory Learning, Defence Science Journal .vol. 67 no. 3, 332-338.

Boukal, Y., Darouach, M., Zasadzinski, M., & Radhy, N. E. (2017). Robust H∞ observer-based control of fractional-order systems with gain parametrization, IEEE Trans. Autom. Control, vol. 62 , no. 11, 5710-5723.

Castañeda, H., Plestan, F., Chriette, A., & León-Morales, J. (2016). Continuous differentiator based on adaptive second-order Sliding-mode control for a 3-DOF Helicopter, IEEE Trans. Ind. Electron., vol. 63, no. 9, 5786-5793

Chakrabarty, A., Fridman, E., Żak, S. H., & Buzzard, G.T. (2018). State and unknown input observers for nonlinear systems with delayed measurements, Automatica, vol. 95, 246–253.

Franklin, G. F., Powell, J. D., & Emami-Naeini, A. (2014). Feedback control of dynamic systems, 7th ed. Pearson Prentice Hall, New Jersey.

Glader, C., Hoegnaes, G., Maekilae, P. M., & Toivonen, H. T. (1991). Approximation of delay system – A case study, Int. J. Control, vol. 53, no. 2, 369-390.

Han, B., Lee, J. S., & Kim, M. (2018). Repetitive controller with phase-lead compensation for Cuk CCM inverter, IEEE Trans. Ind. Electron., vol. 65, no. 3, 2356-2367.

Hansen, J. M., Johansen, T. A., Sokolova, N., & Fossen, T. I. (2019). Nonlinear observer for tightly coupled integrated inertial navigation aided by RTKGNSS measurements. IEEE Trans. Control Syst. Technol., vol. 27, no. 3, 1084-1099.

Ibrir, S. (2004). Linear time-derivative trackers, Automatica, vol. 40, no. 3, 397-405.

Kader, Z., Zheng, G., & Barbot, J. P. (2017). Impulsive observer design for linear systems with delayed outputs, IFAC PapersOnLine 50-1, 1263-1268.

Khalil, H. K. (2017). Cascade high-gain observers in output feedback control, Automatica, vol. 80, 110-118.

Khalil, H. K., & Priess, S. (2016). Analysis of the use of low-pass filters with high-gain observers, IFAC-PapersOnLine, vol. 49, no. 18, 488-492.

Kikuuwe, R., Kanaoka, K., Kumon, T., & Yamamoto, M. (2015). Phase-lead stabilization of force-projecting master-slave systems with a new sliding mode filter, IEEE Trans. Control Syst. Technol., vol. 23, no. 6, 2182-2194.

Levant, A. (1998). Robust exact differentiation via sliding mode technique," Automatica, vol. 34, no. 3, 379–384.

Levant, A. (2003). High-order sliding modes, differentiation and output-feedback control, Int. J. Control, vol. 76, Nos. 9/10, 924-941.

Levant, A., & Livne, M. (2018). Globally convergent differentiators with variable gains, Int. J. Control, vol. 91, 1994-2008.

Levant, A., & Yu, X. (2018). Sliding-mode-based differentiation and filtering, IEEE Trans. Autom. Control, vol. 63,no. 9 , 3061-3067.

Moreno, J. A. (2018). Exact differentiator with varying gains, Int. J. Control, vol. 91, 1983-1993.

Obeid, H., Fridman, L., Laghrouche, S., Harmouche, M., & Golkani, M. A. (2018). Adaptation of Levant's differentiator based on barrier function, Int. J. Control, vol. 91, 2019-2027.

Pekař, L., & Kurečková, E. (2012). Does the higher order mean the better internal delay rational approximation, International Journal of Mathematics and Computers in Simulation, vol. 6, no. 1,

153-160.

Pu, Z., Yuan, R., Yi, J., & Tan, X. (2015). A class of adaptive extended state observers for nonlinear disturbed systems, IEEE Trans. Ind. Electron., vol. 62, no. 9, 5858-5869.

Sanz, R., Garcia, P., Fridman, E., & Albertos, P. (2018). Rejection of mismatched disturbances for systems with input delay via a predictive extended state observer, Int. J. Robust. Nonlinear Control, vol. 28, no. 6, 2457-2467.

Stanković, M. R., Rapaić, M. R., Manojlović, S. M., Mitrović, S. T., Simić, S. M., & Naumović, M. B. (2019). Optimised active disturbance rejection motion control with resonant extended state observer, Int. J. Control, vol. 92, 1815-1826.

Wang, X., Chen, Z., & Yang, G. (2007). Finite-time-convergent differentiator based on singular perturbation technique, IEEE Trans. Autom. Control, vol. 52, no. 9, 1731-1737.

Wang, X., & Lin, H. (2012). Design and frequency analysis of continuous finite-time-convergent differentiator, Aerosp. Sci. Technol., vol. 18, no. 1, 69-78.

Wang, X., & Shirinzadeh, B. (2014). High-order nonlinear differentiator and application to aircraft control, Mechanical Systems and Signal Processing, vol. 46, no. 2, 227-252.

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources, Expert Systems with Applications, vol. 112, no. 12, 258-273.