SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

February 2024, Vol. 67, Iss. 2, 122203:1–122203:13
https://doi.org/10.1007/s11432-022-3815-4

On the convergence of tracking differentiator with multiple stochastic disturbances

Zehao WU[1], Huacheng ZHOU[2*], Baozhu GUO[3] & Feiqi DENG[4]

[1]*School of Mathematics and Big Data, Foshan University, Foshan* 528000*, China;*
[2]*School of Mathematics and Statistics, Central South University, Changsha* 410075*, China;*
[3]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China;*
[4]*Systems Engineering Institute, South China University of Technology, Guangzhou* 510640*, China*

Received 13 June 2022/Revised 6 February 2023/Accepted 25 May 2023/Published online 28 December 2023

Abstract This paper investigates the convergence, noise-tolerance, and filtering performance of a tracking differentiator in the presence of multiple stochastic disturbances for the first time. We consider a general case wherein the input signal is corrupted by additive colored noise, and the tracking differentiator is disturbed by additive colored noise and white noise. The tracking differentiator is shown to track the input signal and its generalized derivatives in the mean square sense. Further, the almost sure convergence can be achieved when the stochastic noise affecting the input signal is vanishing. Herein, numerical simulations are performed to validate the theoretical results.

Keywords tracking differentiator, convergence, noise-tolerance, filtering performance, multiple stochastic disturbances

1 Introduction

It is widely acknowledged that the powerful yet primitive proportional-integral-derivative (PID) control law developed during the 1920s to 1940s dominated control engineering for over a century. Numerous studies on the foundation of PID control have been reported over the past decades, and a remarkable development has recently been made in [1]. However, the derivative control may be impractical because the classical differentiation extraction is sensitive to the noise. With the goal of successfully extracting differential signals from a given input signal, several differentiators have been developed so far. Typical examples include first-order robust exact differentiator (RED) using the sliding mode technique proposed in [2], for which the maximal derivative error is proportional to the square root of the input noise magnitude after a finite-time transient process. The properties and limitations of two different types of linear high-gain differentiators were investigated in [3] in terms of the peaking phenomenon, differentiation error, and noise filtering. A finite-time-convergent differentiator using the singular perturbation technique was designed in [4]. A uniform RED based on the super-twisting algorithm was proposed in [5] and found to be finite-time convergent, with the convergence time being bounded by a constant that is independent of the initial value of the differentiation error. Other recent efforts can be found in [6], which proposes a differentiator/observer with varying exponent gain, and in [7], which proposes a discrete-time differentiation algorithm of arbitrary order based on the uniform RED. However, although the aforementioned studies performed convergence analyses, the input noises were either ignored or assumed to be deterministic.

A noise-tolerant tracking differentiator (TD), which is the first component of the powerful active disturbance rejection control (ADRC) technology [8], was first proposed by Han et al. [9]. The proposed TD not only serves as a differentiation extraction tool but also as a transient profile to ensure that the control plant output can track a smooth signal instead of a jump signal like a step signal in ADRC or PID control, which also exists in the differentiation extraction of the noise-corrupted signals. A comprehensive

* Corresponding author (email: hczhou@amss.ac.cn)

comparison of the commonly used TD and RED was made in the frequency and time domains in [10], and their efficiencies were evaluated by the hardware. The results revealed that the TD outperforms RED in terms of quickness, smoothness, and noise tolerance without causing overshooting. Furthermore, the TD exhibits excellent parameter adaptability and tolerance to signal frequency variations. In addition, compared with the sliding-mode-based differentiators, which typically cause chattering and overshooting, the noise-tolerant TD exhibits favorable smoothness [11, 12]. Finally, the TD effectively reduces peaking value and tolerates measurement noises as compared with the classical high-gain observer-based differentiators [13], and does not require the input signal model in the Kalman filter-based method [14]. To summarize, the efficacy of TD has been validated by numerous numerical experiments and engineering applications, such as high-performance motion control [15, 16], the control of single-axis rotation inertial navigation systems [17], and the speed and position detection system for a maglev train [18].

The convergence of a simple linear TD was first demonstrated in [19] with an application in the online estimation of the frequency of sinusoidal signals. The convergence analysis of the nonlinear and linear TDs for both two-dimensional and high-dimensional cases under weak assumptions was carried out in [20]. The weak convergence of a nonlinear TD based on a finite-time stable system was presented in [21]. A rigorous convergence analysis for the discrete-time optimal control-based TD was presented in [12]. More comprehensive reports on the convergence analysis of linear, nonlinear, and finite-time stable TD can be found in Chapter 2 in [22], where no input noise was considered. To summarize, the convergence of TD in literature has been developed for input signal tracking in transient processes and for the extraction of its generalized derivatives without considering stochastic input noise. However, conventional derivatives tracking was limited to the linear case and did not account for the stochastic input noise. Conventional derivatives tracking is a remarkable ideological breakthrough for practical implementations, and we no longer require the exact value of the derivatives of a signal which is almost always affected by stochastic disturbances. There are numerous studies on stochastic systems in different situations such as those in [23–28]. Motivated by this consideration, we investigate for the first time, the convergence, noise-tolerance, and filtering performance of the TD proposed by Han et al. [9] when the input signal is corrupted by additive colored noise, and the TD is disturbed by additive colored and white noises.

The main contribution and novelty of this study are twofold. First, a rigorous analysis of the TD in the presence of multiple stochastic disturbances is presented, including additive colored and white noises, which evaluates the signal tracking, noise-tolerance, and filtering performances. Second, it is demonstrated that the states of TD track input signal in the transient process as well as its generalized derivatives in the mean square and almost sure sense.

The remaining paper is structured as follows: in Section 2, the problem is formulated, and preliminary information is presented. In Section 3, the main findings are presented. The proofs of the main results are provided in Section 4. The details of numerical simulations are presented in Section 5, followed by concluding remarks in Section 6.

2 Problem formulation and preliminaries

The following notations are used throughout the paper: $\mathbb{E}$ represents the mathematical expectation; $|X|$ represents the absolute value of a scalar X, $\|X\|$ represents the Euclidean norm of a vector X, and $a \wedge b$ is the minimum of real numbers a and b. Further, let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a complete filtered probability space with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geqslant 0}$ on which three mutually independent one-dimensional standard Brownian motions $B_i(t)$ $(i = 1, 2, 3)$ are defined. For notational simplicity, we use $g(t)$ to represent a stochastic process $g(t, \omega)$.

In many cases, stochastic disturbances are modeled by white noise, which is a stationary stochastic process with zero mean and constant spectral density and is the generalized derivative of Brownian motion (see [29, p.51, Theorem 3.14]). However, white noise does not always accurately describe stochastic disturbances in nature because its δ-function correlation is an idealization of the correlations of real processes, which frequently have finite, or even long, correlation times [30]. A more realistic description could be provided by an exponentially correlated process, which is known as colored noise or the Ornstein-Uhlenbeck process [30, 31]. Let $w_i(t)$ $(i = 1, 2)$ represent the colored noise. They are solutions to Itô-type stochastic differential equations (see [30, p.426], [32, p.101]):

$$\mathrm{d}w_i(t) = -\alpha_i w_i(t)\mathrm{d}t + \alpha_i \sqrt{2\beta_i}\mathrm{d}B_i(t), \tag{1}$$

where $\alpha_i > 0$ and β_i are given constants that describe the correlation time and noise intensity, respectively, and the initial values $w_i(0) \in L^2(\Omega;\mathbb{R})$ are independent of $B_i(t)$. In other words, the parameters α_i describe the bandwidth of the noise, while β_i represents its spectral height, and the correlation functions of the processes $w_i(t)$ are more realistic exponential functions but not the δ-ones (see [30]). In the following, α_i and β_i can be unknown constants.

Lemma 1 (The Itô isometry [33])**.** Let $B(t)$ be an m-dimensional standard Brownian motion, and $g : [0,\infty) \times \Omega \to \mathbb{R}^{n\times m}$ is a measurable function satisfying $\mathbb{E}[\int_0^t (g(s))^2 \mathrm{d}s] < \infty$. Then there holds

$$\mathbb{E}\left(\int_0^t g(s)\mathrm{d}B(s)\right)^2 = \mathbb{E}\left[\int_0^t (g(s))^2 \,\mathrm{d}s\right]. \tag{2}$$

Let $v(t)$ represent a time-varying input signal which is supposed to be corrupted by additive colored noise. Accordingly, the input signal is

$$v^*(t) := v(t) + \sigma_1 w_1(t), \tag{3}$$

where σ_1 is a constant that may be unknown and represents the intensity of the colored noise. The stochastic input noise is referred to as vanishing if the noise does not exist, i.e., $\sigma_1 = 0$. Further, we consider the following general case wherein the system constructing TD is disturbed by both additive colored noise and white noise as follows:

$$\begin{cases} \mathrm{d}x_1(t) = x_2(t)\mathrm{d}t, \\ \mathrm{d}x_2(t) = x_3(t)\mathrm{d}t, \\ \vdots \\ \mathrm{d}x_{n-1}(t) = x_n(t)\mathrm{d}t, \\ \mathrm{d}x_n(t) = r^n f\left(x_1(t) - v^*(t), \frac{x_2(t)}{r}, \ldots, \frac{x_n(t)}{r^{n-1}}\right)\mathrm{d}t + \sigma_2 w_2(t)\mathrm{d}t + \sigma_3 \mathrm{d}B_3(t), \end{cases} \tag{4}$$

where $r > 0$ is a tuning parameter, $f : \mathbb{R}^n \to \mathbb{R}$ is a suitable known function chosen to satisfy Assumption 1, $\sigma_2 w_2(t) + \sigma_3 \dot{B}_3(t)$ represents the multiple stochastic disturbances with σ_i $(i = 2, 3)$ being constants that may be unknown, and $\dot{B}_3(t)$ is the white noise which is the formal derivative of the Brownian motion. Next, we define $x(t) = (x_1(t), \ldots, x_n(t))$. The TD proposed by Han et al. [9] is a special case of (4) with $\sigma_i = 0, i = 1, 2, 3$. The evaluation of such a TD is based on three factors. First, such a TD in a noisy environment is quite general, whereas a TD with no noise corruption is simply a special case of $\sigma_2 = \sigma_3 = 0$. Second, quantization errors caused by the digital implementation of a TD are unavoidable and can be regarded as process noise. Finally, TD is the first component of the powerful ADRC that has been hardwired into general purpose control chips manufactured by industry giants such as Texas Instruments [34], where the hardware may operate in a noisy environment.

In Han's original formulation [8, 9], the convergence of a TD (4) (without noises for $v(t)$) means that

$$\int_0^T |x_1(t) - v(t)|\mathrm{d}t \to 0 \text{ as } r \to \infty$$

for any $T > 0$, and $x_i(t)$ $(i = 2, \ldots, n)$ are regarded as generalized derivatives of $v(t)$ even when the classical derivatives $v^{(i-1)}(t)$ do not exist. This is the first notable point of the Han's TD. Second, Han's TD serves as a transient profile, allowing the output of the control plant to track a smooth signal $x_1(t)$ rather than $v(t)$ like jump step signal in ADRC or PID control.

This paper considers the convergence of the tracking error $x_1(t) - v(t)$ in the mean square and almost sure sense. Specifically, we are concerned about the result that $x_1(t)$ tracks $v(t)$ uniformly in $[T, \infty)$ for any $T > 0$, which takes the transient process into account. The noise tolerance and filtering performance of TD are also considered. In this sense, the TD tracks the generalized derivatives of the input signal in both mean square and almost sure sense, regardless of whether the higher classical derivatives exist. In conventional research, the convergence of $x_i(t) - v^{(i-1)}(t)$ for $i = 2, \ldots, n$ is also estimated, but in our study, this is only possible for a linear TD without noise, as discussed in [20]. From a theoretical standpoint, the convergence of $x_i(t) - v^{(i-1)}(t)$ for $i = 2, \ldots, n$ is indeed an important and challenging problem that should be investigated further whenever classical derivatives $v^{(i-1)}(t)$ do exist.

Since the solution of (4) is dependent on the tuning parameter r, hereafter, we drop r from solutions for notational simplicity.

The following Assumption 1 is a prior assumption about the known function $f(\cdot)$ used in the construction of TD (4).

Assumption 1. The $f:\mathbb{R}^n\to\mathbb{R}$ is a locally Lipschitz continuous function regarding its arguments, $f(0,\ldots,0)=0$, and there exist known constants $\lambda_i>0$ $(i=1,2,3,4)$ and a twice continuously differentiable function $V:\mathbb{R}^n\to[0,\infty)$, which is positive definite and radially unbounded such that

$$\begin{aligned}
&\lambda_1\|z\|^2\leqslant V(z)\leqslant\lambda_2\|z\|^2,\ \lambda_3\|z\|^2\leqslant W(z)\leqslant\lambda_4\|z\|^2,\\
&\sum_{i=1}^{n-1}\frac{\partial V(z)}{\partial z_i}z_{i+1}+\frac{\partial V(z)}{\partial z_n}f(z)\leqslant -W(z),\\
&\left|\frac{\partial V(z)}{\partial z_j}\right|\leqslant c_1\|z\|,\ \left|\frac{\partial^2 V(z)}{\partial z_j^2}\right|\leqslant c_2,\\
&\forall\ z=(z_1,z_2,\ldots,z_n)\in\mathbb{R}^n,\ j=1,n,
\end{aligned}\tag{5}$$

for some non-negative continuous function $W:\mathbb{R}^n\to[0,\infty)$ and constants $c_i>0$ $(i=1,2)$.

Remark 1. In general, Assumption 1 ensures that the function $f:\mathbb{R}^n\to\mathbb{R}$ is chosen such that the zero equilibrium state of the system

$$\dot{z}(t)=(z_2(t),z_3(t),\ldots,f(z(t)))\tag{6}$$

is globally exponentially stable with $z=(z_1,z_2,\ldots,z_n)$. The linear TD with f being a linear function is a special case of the TD (4).

The solution of (1) can be expressed as

$$w_i(t)=\mathrm{e}^{-\alpha_i t}w_i(0)+\int_0^t \mathrm{e}^{-\alpha_i(t-s)}\alpha_i\sqrt{2\beta_i}\mathrm{d}B_i(s).\tag{7}$$

Let

$$\gamma_i=\mathbb{E}|w_i(0)|^2+\alpha_i\beta_i,\ i=1,2.\tag{8}$$

The Itô isometric formula in Lemma 1 makes it simple to demonstrate that the second moments of $w_i(t)$ $(i=1,2)$ are bounded,

$$\begin{aligned}
\mathbb{E}|w_i(t)|^2&=\mathrm{e}^{-2\alpha_i t}\mathbb{E}|w_i(0)|^2+\mathbb{E}\left|\int_0^t \mathrm{e}^{-\alpha_i(t-s)}\alpha_i\sqrt{2\beta_i}\mathrm{d}B_i(s)\right|^2\\
&\leqslant\mathbb{E}|w_i(0)|^2+2\alpha_i^2\beta_i\int_0^t \mathrm{e}^{-2\alpha_i(t-s)}\mathrm{d}s\leqslant\gamma_i,\ \forall t\geqslant 0.
\end{aligned}\tag{9}$$

This is why the TD may be feasible when the input signal is disturbed by additive colored noise.

3 Main results

Set $\hat{B}_1(t)=\sqrt{r}B_1\left(\frac{t}{r}\right),\ \hat{B}_3(t)=\sqrt{r}B_3\left(\frac{t}{r}\right)$. Note that for any $r>0$, $\hat{B}_1(t)$ and $\hat{B}_3(t)$ remain mutually independent one-dimensional standard Brownian motions. According to the definition of $v^*(t)$ in (3), it follows that

$$\mathrm{d}v^*(t)=\dot{v}(t)\mathrm{d}t-\sigma_1\alpha_1 w_1(t)\mathrm{d}t+\sigma_1\alpha_1\sqrt{2\beta_1}\mathrm{d}B_1(t),\tag{10}$$

and then

$$\mathrm{d}v^*\left(\frac{t}{r}\right)=\dot{v}\left(\frac{t}{r}\right)\mathrm{d}t-\frac{\sigma_1\alpha_1}{r}w_1\left(\frac{t}{r}\right)\mathrm{d}t+\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\mathrm{d}\hat{B}_1(t),\tag{11}$$

where $\dot{v}(\frac{t}{r})$ denotes, subsequently, the derivative of $v(\frac{t}{r})$ with respect to the time t. For $i=2,\ldots,n$, set

$$y_1(t)=x_1\left(\frac{t}{r}\right)-v^*\left(\frac{t}{r}\right),\ \ y_i(t)=\frac{1}{r^{i-1}}x_i\left(\frac{t}{r}\right).\tag{12}$$

It is worth noting that the transformation is one-to-one because we have

$$x_1(t) = y_1(rt) + v^*(t), \;\; x_i(t) = r^{i-1}y_i(rt), \; i = 2, \ldots, n. \tag{13}$$

A direct computation reveals that $y(t) = (y_1(t), \ldots, y_n(t))$ satisfies the following Itô-type stochastic differential equation:

$$\begin{cases} \mathrm{d}y_1(t) = y_2(t)\mathrm{d}t - \dot{v}\left(\frac{t}{r}\right)\mathrm{d}t + \frac{\sigma_1\alpha_1}{r}w_1\left(\frac{t}{r}\right)\mathrm{d}t - \frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\mathrm{d}\hat{B}_1(t), \\ \mathrm{d}y_2(t) = y_3(t)\mathrm{d}t, \\ \vdots \\ \mathrm{d}y_{n-1}(t) = y_n(t)\mathrm{d}t, \\ \mathrm{d}y_n(t) = f(y(t))\mathrm{d}t + \frac{\sigma_2}{r^n}w_2\left(\frac{t}{r}\right)\mathrm{d}t + \frac{\sigma_3}{r^{n-\frac{1}{2}}}\mathrm{d}\hat{B}_3(t). \end{cases} \tag{14}$$

Set

$$\begin{cases} N_1 = 2\mathbb{E}V(y(0)) + \frac{32c_1^2\sigma_1^2\alpha_1^2\beta_1}{r}, \\ N_2 = \frac{c_1^2M^2}{r} + \frac{c_1^2\sigma_1^2\alpha_1^2\gamma_1}{r} + \frac{2c_2\sigma_1^2\alpha_1^2\beta_1}{r} + \frac{c_1^2\sigma_2^2\gamma_2}{r^n} + \frac{c_2\sigma_3^2}{r^{2n-1}} + \frac{32c_1^2\sigma_3^2}{r^{2n-1}\lambda_1}, \\ N_3 = \frac{2}{\lambda_1 r} + \frac{1}{\lambda_1 r^n} + \frac{2}{\lambda_1}, \end{cases} \tag{15}$$

where $M := \sup_{t\geqslant 0}(|v(t)| + |\dot{v}(t)|) < \infty$ as assumed in Lemma 2.

We first present Lemma 2 to demonstrate the existence and uniqueness of the global solution for the system (14) and provide an estimate of the second moment of the global solution.

Lemma 2. Suppose that $v : [0,\infty) \to \mathbb{R}$ is a continuously differentiable function satisfying $M := \sup_{t\geqslant 0}(|v(t)| + |\dot{v}(t)|) < \infty$, Assumption 1 holds, and the tuning parameter r is chosen so that $r \geqslant 1$. Then for any initial value $x(0) \in \mathbb{R}^n$, system (14) admits a unique global solution satisfying

$$\mathbb{E}\left(\sup_{0\leqslant s\leqslant t}\|y(s)\|^2\right) \leqslant \frac{1}{\lambda_1}\left(N_1 + \frac{N_2}{N_3}\right)e^{N_3 t}, \; \forall t \geqslant 0, \tag{16}$$

where the constants N_i $(i = 1, 2, 3)$ are defined in (15).

Proof. See “Proof of Lemma 2” in Section 4.

In what follows, a value range of the tuning parameter to ensure the convergence of TD (4) can be specified as

$$R_0 := \left\{ r \geqslant 1 : \frac{1}{r} + \frac{1}{2r^{2n-1}} \leqslant \frac{\theta\lambda_3}{\lambda_2} \right\}, \tag{17}$$

where $\theta \in (0, 1)$ is any chosen parameter. For any positive constants $\theta, \lambda_2, \lambda_3$, it can be obtained that $[\frac{2\lambda_2}{\theta\lambda_3}, \infty) \subseteq R_0$, implying that R_0 is never empty. Furthermore, note that as $\theta \in (0, 1)$ increases, the range R_0 increases as well, but the exponential decay rate $\frac{(1-\theta)\lambda_3}{\lambda_2}$ associated with the tracking error decreases.

The convergence result of TD (4) in the presence of multiple stochastic disturbances is presented as the following Theorem 1.

Theorem 1. Assume that $v : [0,\infty) \to \mathbb{R}$ is a continuously differentiable function satisfying $M := \sup_{t\geqslant 0}(|v(t)| + |\dot{v}(t)|) < \infty$ and Assumption 1 holds. Then, for any tuning parameter $r \in R_0$, initial value $x(0) \in \mathbb{R}^n$, and $T > 0$, the TD (4) admits a unique global solution satisfying

$$\mathbb{E}|x_1(t) - v(t)|^2 \leqslant \frac{(1 + \frac{1}{\mu})\Gamma}{r} + (1 + \mu)\sigma_1^2\gamma_1 \tag{18}$$

uniformly in $t \in [T, \infty)$, where μ is any positive constant and Γ is a positive constant independent of r;

$$\limsup_{r\to\infty} \mathbb{E}|x_1(t) - v(t)|^2 \leqslant \sigma_1^2\gamma_1 \tag{19}$$

uniformly in $t \in [T, \infty)$;

$$\lim_{r\to\infty} |x_1(t) - v(t)| = 0 \;\text{ almost surely} \tag{20}$$

uniformly in $t \in [T, \infty)$ when $\sigma_1 = 0$.

Proof. See "Proof of Theorem 1" in Section 4.

Remark 2. We can see from (18) and (19) that the upper bound of the tracking error in the mean square can approach $\sigma_1^2\gamma_1$ arbitrarily and quickly by tuning the parameter r to be sufficiently large because the convergence time T is any positive constant. The theoretical results also show noise-tolerance and filtering performance because the input noise (which could be some kinds of measurement noise) is not amplified in the upper bound and the process noises "$\sigma_2 w_2(t)+\sigma_3\dot{B}_3(t)$" are filtered completely. It seems impossible to make the tracking error as small as possible when $\sigma_1 \neq 0$ in (3), which is similar to the input-to-state stability of stochastic systems with nonvanishing noise vector fields [35, 36].

Remark 3. It is worth noting that the choice of the function $f(\cdot)$ ensures that the "nominal part" of the tracking error system (14) defined in (6) is exponentially stable with the decay rate $\frac{\lambda_3}{\lambda_2}$ which depends on $f(\cdot)$. According to the definition of Γ from (40) in Section 4, the constant Γ, which is a part of the tracking error (depending on $f(\cdot)$), becomes smaller as the decay rate $\frac{\lambda_3}{\lambda_2}$ increases. Another advantage that could be mentioned is that as the decay rate $\frac{\lambda_3}{\lambda_2}$ increases, so does the value range of the tuning parameter r defined in (17).

Finally, we indicate an important fact that $x_i(t)$ $(i = 2, 3, \ldots, n)$ can always be considered an approximation of the corresponding $(i-1)$-th derivative of $v(t)$ in terms of generalized derivative regardless of whether classical derivatives of $v(t)$ exist. In fact, for any $a > 0$, let $C_0^\infty(0, a)$ be the set containing all infinitely differentiable functions with compact support on $(0, a)$. Remember that for any locally integrable function $h : (0, a) \to \mathbb{R}$, the usual $(i-1)$-th generalized derivative of h, still denoted by $h^{(i-1)}$, always exists in the sense of distribution defined as a functional on $C_0^\infty(0, a)$ that

$$h^{(i-1)}(\varphi) = (-1)^{i-1}\int_0^a h(t)\varphi^{(i-1)}(t)\mathrm{d}t, \tag{21}$$

for every test function $\varphi \in C_0^\infty(0, a)$ and $2 \leqslant i \leqslant n$ (see [22, p.43]). Furthermore, a generalized stochastic process Φ is simply a random generalized function in the following sense: for each test function $\varphi \in C_0^\infty(0, a)$, a random variable $\Phi(\varphi)$ is assigned so that the functional Φ on $C_0^\infty(0, a)$ is linear and continuous (see [29, p.50]). Thus, for each $i = 2, 3, \ldots, n$, x_i can be considered a generalized stochastic process in the sense that

$$x_i(\varphi) = \int_0^a x_i(t)\varphi(t)\mathrm{d}t,\ \forall\varphi \in C_0^\infty(0, a). \tag{22}$$

For each $i = 2, 3, \ldots, n$, the state x_i of the TD (4) is convergent to the $(i-1)$-th generalized derivative of the input signal v in the mean square and almost sure sense, as summarized in Theorem 2.

Theorem 2. Assume that $v : [0, a] \to \mathbb{R}$ is continuously differentiable and that Assumption 1 holds. Then, for any initial value $x(0) \in \mathbb{R}^n$, the TD (4) admits a unique global solution, and for any $\varphi \in C_0^\infty(0, a)$ and all $i = 2, 3, \ldots, n$, there holds

$$\limsup_{r\to\infty} \mathbb{E}|x_i(\varphi) - v^{(i-1)}(\varphi)|^2 \leqslant a^2 \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2\sigma_1^2\gamma_1, \tag{23}$$

$$\lim_{r\to\infty} |x_i(\varphi) - v^{(i-1)}(\varphi)| = 0 \ \text{ almost surely} \tag{24}$$

when the additive colored noise affecting the input signal is vanishing, i.e., $\sigma_1 = 0$.

Proof. See " Proof of Theorem 2" in Section 4.

4 Proofs of main results

In this section, we give proofs of the main results.

Proof of Lemma 2. By (1), $w_i(t)$ $(i = 1, 2)$ can be regarded as the augmented state variables of the system (14). Because the function $f(\cdot)$ satisfies the local Lipschitz condition, it follows from the existence-and-unique theorem for Itô-type stochastic systems (see [32, p.58,Theorem 3.6]) that there exists a unique maximal local solution $y(t)$ over $t \in [0, \tau)$ where τ is the explosion time. To show that $y(t)$ is a global solution, we simply need to show that $\tau = \infty$ almost surely. For each integer $m \geqslant 1$, define the stopping time $\tau_m = \tau \wedge \inf\{t : 0 \leqslant t < \tau, \|y(t)\| \geqslant m\}$, and set $\inf \emptyset = \infty$. Since $\{t : 0 \leqslant t < \tau, \|y(t)\| \geqslant m+1\} \subset \{t : 0 \leqslant t < \tau, \|y(t)\| \geqslant m\}$, we have $\tau_m \uparrow \tau$ almost surely as $m \to \infty$. According to the Itô's formula,

$$\begin{aligned} V(y(t)) &= V(y(0)) + \int_0^t \left[\sum_{i=1}^{n-1} \frac{\partial V(y(s))}{\partial y_i} y_{i+1}(s) + \frac{\partial V(y(s))}{\partial y_n} f(y(s))\right] \mathrm{d}s \\ &\quad + \int_0^t \left[\frac{\partial V(y(s))}{\partial y_1}\left(-\frac{1}{r}\frac{\mathrm{d}v(u)}{\mathrm{d}u}\Big|_{u=\frac{s}{r}} + \frac{\sigma_1\alpha_1}{r} w_1\left(\frac{s}{r}\right)\right)\right] \mathrm{d}s \\ &\quad + \frac{\sigma_1^2\alpha_1^2\beta_1}{r}\int_0^t \frac{\partial^2 V(y(s))}{\partial y_1^2}\mathrm{d}s + \frac{\sigma_2}{r^n}\int_0^t \frac{\partial V(y(s))}{\partial y_n} w_2\left(\frac{s}{r}\right)\mathrm{d}s + \frac{\sigma_3^2}{2r^{2n-1}}\int_0^t \frac{\partial^2 V(y(s))}{\partial y_n^2}\mathrm{d}s \\ &\quad - \int_0^t \frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s) + \int_0^t \frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s). \end{aligned} \tag{25}$$

Consequently, from Assumption 1, (9), and Young's inequality, it follows that

$$\begin{aligned} &\mathbb{E}\left(\sup_{0\leqslant u\leqslant t} V(y(u\wedge\tau_m))\right) \\ &\leqslant \mathbb{E}V(y(0)) + \mathbb{E}\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m}\left[\frac{1}{2\lambda_1 r}V(y(s)) + \frac{c_1^2M^2}{2r} + \frac{1}{2\lambda_1 r}V(y(s)) + \frac{c_1^2\sigma_1^2\alpha_1^2}{2r}\left|w_1\left(\frac{s}{r}\right)\right|^2\right. \\ &\quad \left. + \frac{c_2\sigma_1^2\alpha_1^2\beta_1}{r} + \frac{1}{2\lambda_1 r^n}V(y(s)) + \frac{c_1^2\sigma_2^2}{2r^n}\left|w_2\left(\frac{s}{r}\right)\right|^2 + \frac{c_2\sigma_3^2}{2r^{2n-1}}\right]\mathrm{d}s \\ &\quad + \mathbb{E}\left(\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m} -\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s)\right) + \mathbb{E}\left(\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m}\frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s)\right) \\ &\leqslant \mathbb{E}V(y(0)) + \int_0^t\left[\frac{c_1^2M^2}{2r} + \frac{c_1^2\sigma_1^2\alpha_1^2\gamma_1}{2r} + \frac{c_2\sigma_1^2\alpha_1^2\beta_1}{r} + \frac{c_1^2\sigma_2^2\gamma_2}{2r^n} + \frac{c_2\sigma_3^2}{2r^{2n-1}}\right. \\ &\quad \left. + \left(\frac{1}{\lambda_1 r} + \frac{1}{2\lambda_1 r^n}\right)\mathbb{E}V(y(s\wedge\tau_m))\right]\mathrm{d}s + \mathbb{E}\left(\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m} -\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s)\right) \\ &\quad + \mathbb{E}\left(\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m}\frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s)\right). \end{aligned} \tag{26}$$

By Assumption 1, it is seen that $\int_0^{t\wedge\tau_m} -\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s)$ is a martingale on $t \geqslant 0$, and so is for $\int_0^{t\wedge\tau_m}\frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s)$. By the Burkholder-Davis-Gundy inequality ([32, Theorem 1.7.3]),

$$\begin{aligned} &\mathbb{E}\left(\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m} -\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s)\right) \\ &\quad \leqslant 4\sqrt{2}\mathbb{E}\left(\int_0^t\left|\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s\wedge\tau_m))}{\partial y_1}\right|^2\mathrm{d}s\right)^{\frac{1}{2}} \\ &\quad \leqslant \frac{16c_1^2\sigma_1^2\alpha_1^2\beta_1}{r} + \frac{1}{\lambda_1}\int_0^t\mathbb{E}V(y(s\wedge\tau_m))\mathrm{d}s, \end{aligned} \tag{27}$$

and

$$\begin{aligned} &\mathbb{E}\left(\sup_{0\leqslant u\leqslant t}\int_0^{u\wedge\tau_m}\frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s)\right) \\ &\quad \leqslant \frac{4\sqrt{2}\sigma_3}{r^{n-\frac{1}{2}}}\mathbb{E}\left(\int_0^t\left|\frac{\partial V(y(s\wedge\tau_m))}{\partial y_n}\right|^2\mathrm{d}s\right)^{\frac{1}{2}} \\ &\quad \leqslant \frac{4\sqrt{2}\sigma_3 c_1}{r^{n-\frac{1}{2}}\sqrt{\lambda_1}}\mathbb{E}\left(\sup_{0\leqslant s\leqslant t}V(y(s\wedge\tau_m))\int_0^t 1\mathrm{d}s\right)^{\frac{1}{2}} \\ &\quad \leqslant \frac{1}{2}\mathbb{E}\left(\sup_{0\leqslant s\leqslant t}V(y(s\wedge\tau_m))\right) + \int_0^t\frac{16c_1^2\sigma_3^2}{r^{2n-1}\lambda_1}\mathrm{d}s. \end{aligned} \tag{28}$$

By combining (26)–(28), we acquire

$$\mathbb{E}\left(\sup_{0\leqslant s\leqslant t}V(y(s\wedge\tau_m))\right)$$

$$\begin{aligned}&\leqslant N_1+\int_0^t (N_2+N_3\mathbb{E}V(y(s\wedge\tau_m)))\mathrm{d}s\\&=N_1+N_3\int_0^t\left(\frac{N_2}{N_3}+\mathbb{E}\sup_{0\leqslant u\leqslant s}V(y(u\wedge\tau_m))\right)\mathrm{d}s,\end{aligned}\tag{29}$$

where N_i $(i=1,2,3)$ are specified in (15). Now, using Gronwall's inequality ([32, Theorem 1.8.1]) yields

$$\frac{N_2}{N_3}+\mathbb{E}\left(\sup_{0\leqslant s\leqslant t}V(y(s\wedge\tau_m))\right)\leqslant\left(N_1+\frac{N_2}{N_3}\right)\mathrm{e}^{N_3t}.\tag{30}$$

Thus,

$$\mathbb{E}\left(\sup_{0\leqslant s\leqslant t\wedge\tau_m}\|y(s)\|^2\right)\leqslant\frac{1}{\lambda_1}\mathbb{E}\left(\sup_{0\leqslant s\leqslant t\wedge\tau_m}V(y(s))\right)\leqslant\frac{1}{\lambda_1}\left(N_1+\frac{N_2}{N_3}\right)\mathrm{e}^{N_3t}.\tag{31}$$

This implies that

$$m^2P\{\tau_m\leqslant t\}\leqslant\frac{1}{\lambda_1}\left(N_1+\frac{N_2}{N_3}\right)\mathrm{e}^{N_3t}.\tag{32}$$

Passing to the limit as $m\to\infty$ yields $P\{\tau\leqslant t\}=0$ which in turn yields $P\{\tau>t\}=1$. Because $t\geqslant 0$ is arbitrary, we then have $\tau=\infty$ almost surely, and $y(t)$ exists globally. Furthermore, passing to the limit as $m\to\infty$ for (31) again yields (16) from Fatou's Lemma. This completes the proof of the Lemma 2.

Proof of Theorem 1. According to Lemma 2, the TD (4) admits a unique global solution and $\int_0^t\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s)$ is a martingale on $t\geqslant 0$, and so is for $\int_0^t\frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s)$. Thus, for all $t\geqslant 0$, it follows that

$$\mathbb{E}\int_0^t\frac{\sigma_1\alpha_1\sqrt{2\beta_1}}{\sqrt{r}}\frac{\partial V(y(s))}{\partial y_1}\mathrm{d}\hat{B}_1(s)=0,\ \ \mathbb{E}\int_0^t\frac{\sigma_3}{r^{n-\frac{1}{2}}}\frac{\partial V(y(s))}{\partial y_n}\mathrm{d}\hat{B}_3(s)=0.\tag{33}$$

Finding the derivative of $\mathbb{E}V(y(t))$ with respect to t, it follows from Assumption 1, (9), (25), (33), $r\in\mathbb{R}_0$, and Young's inequality that

$$\begin{aligned}\frac{\mathrm{d}\mathbb{E}V(y(t))}{\mathrm{d}t}&=\mathbb{E}\left(\sum_{i=1}^{n-1}\frac{\partial V(y(t))}{\partial y_i}y_{i+1}(t)+\frac{\partial V(y(t))}{\partial y_n}f(y(t))\right)\\&\quad+\mathbb{E}\left(\frac{\partial V(y(t))}{\partial y_1}\left(-\frac{1}{r}\frac{\mathrm{d}v(u)}{\mathrm{d}u}\bigg|_{u=\frac{t}{r}}+\frac{\sigma_1\alpha_1}{r}w_1\left(\frac{t}{r}\right)\right)\right)\\&\quad+\mathbb{E}\left(\frac{\sigma_1^2\alpha_1^2\beta_1}{r}\frac{\partial^2V(y(t))}{\partial y_1^2}+\frac{\sigma_2}{r^n}\frac{\partial V(y(t))}{\partial y_n}w_2\left(\frac{t}{r}\right)\right)+\frac{\sigma_3^2}{2r^{2n-1}}\mathbb{E}\frac{\partial^2V(y(t))}{\partial y_n^2}\\&\leqslant-\mathbb{E}W(y(t))+\frac{\lambda_1}{2r}\mathbb{E}\|y(t)\|^2+\frac{c_1^2M^2}{2\lambda_1r}+\frac{\lambda_1}{2r}\mathbb{E}\|y(t)\|^2+\frac{c_1^2\sigma_1^2\alpha_1^2}{2\lambda_1r}\mathbb{E}\left|w_1\left(\frac{t}{r}\right)\right|^2\\&\quad+\frac{c_2\sigma_1^2\alpha_1^2\beta_1}{r}+\frac{\lambda_1}{2r^{2n-1}}\mathbb{E}\|y(t)\|^2+\frac{c_1^2\sigma_2^2}{2\lambda_1r}\mathbb{E}\left|w_2\left(\frac{t}{r}\right)\right|^2+\frac{c_2\sigma_3^2}{2r^{2n-1}}\\&\leqslant-\frac{\lambda_3}{\lambda_2}\mathbb{E}V(y(t))+\frac{1}{2r}\mathbb{E}V(y(t))+\frac{c_1^2M^2}{2\lambda_1r}+\frac{1}{2r}\mathbb{E}V(y(t))+\frac{c_1^2\sigma_1^2\alpha_1^2\gamma_1}{2\lambda_1r}\\&\quad+\frac{c_2\sigma_1^2\alpha_1^2\beta_1}{r}+\frac{1}{2r^{2n-1}}\mathbb{E}V(y(t))+\frac{c_1^2\sigma_2^2\gamma_2}{2\lambda_1r}+\frac{c_2\sigma_3^2}{2r^{2n-1}}\\&\leqslant-\frac{(1-\theta)\lambda_3}{\lambda_2}\mathbb{E}V(y(t))+\frac{\Gamma_1}{r},\end{aligned}\tag{34}$$

where

$$\Gamma_1:=\frac{c_1^2M^2}{2\lambda_1}+\frac{c_1^2\sigma_1^2\alpha_1^2\gamma_1}{2\lambda_1}+c_2\sigma_1^2\alpha_1^2\beta_1+\frac{c_1^2\sigma_2^2\gamma_2}{2\lambda_1}+\frac{c_2\sigma_3^2}{2},\tag{35}$$

and $\theta\in(0,1)$ is given in (17). This, combining with Assumption 1, yields

$$\begin{aligned}\mathbb{E}V(y(t))&\leqslant\mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2}t}\mathbb{E}V(y(0))+\frac{\Gamma_1}{r}\int_0^t\mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2}(t-s)}\mathrm{d}s\\&\leqslant\lambda_2\mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2}t}\mathbb{E}\|y(0)\|^2+\frac{\lambda_2\Gamma_1}{\lambda_3(1-\theta)r}.\end{aligned}\tag{36}$$

Given that

$$\mathbb{E}\|y(0)\|^2 = \mathbb{E}|x_1(0) - v(0) - \sigma_1 w_1(0)|^2 + \sum_{i=1}^{n-1} \frac{1}{r^{2i}} |x_{i+1}(0)|^2, \tag{37}$$

it can be deduced that for any $T > 0$, there is a positive constant

$$\Gamma_2 := \sup_{r \in R_0} \left(\mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2} rT} r \right) \cdot \left[\mathbb{E}|x_1(0) - v(0) - \sigma_1 w_1(0)|^2 + \sum_{i=1}^{n-1} |x_{i+1}(0)|^2 \right] \tag{38}$$

which is independent of r. This is because $h(r) := \mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2} rT} r$ is continuous with respect to r. Because $\lim_{r \to \infty} h(r) = 0$, $h(r)$ is bounded on the domain R_0, i.e., there is a positive constant N independent of r such that $N = \sup_{r \in R_0} (\mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2} rT} r)$. As a result, Γ_2 is independent of r and so $\mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2} rT} \mathbb{E}\|y(0)\|^2 \leqslant \frac{\Gamma_2}{r}$. Therefore, for all $t \in [T, \infty)$,

$$\begin{aligned} \mathbb{E}|x_1(t) - v^*(t)|^2 &= \mathbb{E}|y_1(rt)|^2 \leqslant \mathbb{E}\|y(rt)\|^2 \leqslant \frac{1}{\lambda_1} \mathbb{E}V(y(rt)) \\ &\leqslant \frac{\lambda_2}{\lambda_1} \mathrm{e}^{-\frac{(1-\theta)\lambda_3}{\lambda_2} rT} \mathbb{E}\|y(0)\|^2 + \frac{\lambda_2 \Gamma_1}{\lambda_3 (1-\theta) \lambda_1 r} \leqslant \frac{\Gamma}{r}, \end{aligned} \tag{39}$$

where

$$\Gamma := \frac{\lambda_2 \Gamma_2}{\lambda_1} + \frac{\lambda_2 \Gamma_1}{\lambda_3 (1-\theta) \lambda_1} \tag{40}$$

is a positive constant independent of r. Using the inequality $(a+b)^2 \leqslant (1+\frac{1}{\mu})a^2 + (1+\mu)b^2$ for any $\mu > 0$ and $a, b \in \mathbb{R}$, it is shown by (9) and (39) that

$$\begin{aligned} \mathbb{E}|x_1(t) - v(t)|^2 &\leqslant \left(1 + \frac{1}{\mu}\right) \mathbb{E}|x_1(t) - v^*(t)|^2 + (1+\mu)\sigma_1^2 \mathbb{E}|w_1(t)|^2 \\ &\leqslant \frac{(1+\frac{1}{\mu})\Gamma}{r} + (1+\mu)\sigma_1^2 \gamma_1 \end{aligned} \tag{41}$$

uniformly in $t \in [T, \infty)$. Given that $\mu > 0$ is arbitrary, it follows from (41) that

$$\limsup_{r \to \infty} \mathbb{E}|x_1(t) - v(t)|^2 \leqslant \sigma_1^2 \gamma_1 \tag{42}$$

uniformly in $t \in [T, \infty)$.

Furthermore, when $\sigma_1 = 0$, it follows from (42) that

$$\lim_{r \to \infty} \mathbb{E}|x_1(t) - v(t)|^2 = 0 \tag{43}$$

uniformly in $t \in [T, \infty)$. Thus, for any $\varepsilon := \frac{1}{m^4} > 0, m \in \mathbb{N}^+$, there exists an m-dependent constant $r^* = r^*(m)$ such that

$$\mathbb{E}|x_1(t) - v(t)|^2 < \frac{1}{m^4} \tag{44}$$

uniformly in $t \in [T, \infty)$ for all $r \geqslant r^*$. By Chebyshev's inequality ([32, p.5]), it has

$$P\left\{\omega : |x_1(t) - v(t)| > \frac{1}{m}\right\} \leqslant \frac{1}{m^2} \tag{45}$$

uniformly in $t \in [T, \infty)$ for all $r \geqslant r^*$. By Borel-Cantelli's lemma ([32, p.7]), it can also be obtained that for almost all $\omega \in \Omega$, there exists an $m_0 = m_0(\omega)$ such that

$$|x_1(t) - v(t)| \leqslant \frac{1}{m} \tag{46}$$

uniformly in $t \in [T, \infty)$ whenever $m \geqslant m_0, r \geqslant r^*$. Therefore, for almost all $\omega \in \Omega$,

$$\limsup_{r\to\infty} |x_1(t) - v(t)| \leqslant \frac{1}{m} \tag{47}$$

whenever $m \geqslant m_0$. Setting $m \to \infty$ gives

$$\lim_{r\to\infty} |x_1(t) - v(t)| = 0 \text{ almost surely} \tag{48}$$

uniformly in $t \in [T, \infty)$ when $\sigma_1 = 0$. This completes the proof of the Theorem 1.

Proof of Theorem 2. Through (22) and performing the integration by parts, it can be easily obtained that for each $i = 2, 3, \ldots, n$,

$$x_i(\varphi) = (-1)^{(i-1)} \int_0^a x_1(t)\varphi^{(i-1)}(t)\mathrm{d}t,\ \forall \varphi \in C_0^\infty(0, a). \tag{49}$$

According to Theorem 1, (49), and the definition of the generalized derivative in (21), for each $i = 2, 3, \ldots, n$ and any $0 < \xi < a$, it follows that

$$\begin{aligned}
\mathbb{E}|x_i(\varphi) - v^{(i-1)}(\varphi)|^2 &= \mathbb{E}\left|\int_0^a x_i(t)\varphi(t)\mathrm{d}t - (-1)^{(i-1)}\int_0^a v(t)\varphi^{(i-1)}(t)\mathrm{d}t\right|^2 \\
&= \mathbb{E}\left|\int_0^a (x_1(t) - v(t))\varphi^{(i-1)}(t)\mathrm{d}t\right|^2 \\
&\leqslant a\int_0^a \mathbb{E}|x_1(t) - v(t)|^2\mathrm{d}t \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2 \\
&\leqslant a\int_0^\xi \mathbb{E}|x_1(t) - v(t)|^2\mathrm{d}t \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2 \\
&\quad + a\int_\xi^a \mathbb{E}|x_1(t) - v(t)|^2\mathrm{d}t \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2 \\
&\leqslant \xi a \max_{0\leqslant t\leqslant \xi} \mathbb{E}|x_1(t) - v(t)|^2 \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2 \\
&\quad + a(a-\xi)\sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2 \left(\left(1 + \frac{1}{\mu}\right)\frac{\Gamma}{r} + (1+\mu)\sigma_1^2\gamma_1\right).
\end{aligned} \tag{50}$$

Because $\mu > 0$ is arbitrary, passing to the limit as $r \to \infty$ yields

$$\begin{aligned}
&\limsup_{r\to\infty} \mathbb{E}|x_i(\varphi) - v^{(i-1)}(\varphi)|^2 \\
&\quad \leqslant \xi a \max_{0\leqslant t\leqslant \xi} \mathbb{E}|x_1(t) - v(t)|^2 \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2 + a(a-\xi)\sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2\sigma_1^2\gamma_1.
\end{aligned} \tag{51}$$

Setting $\xi \to 0$, we then acquire

$$\limsup_{r\to\infty} \mathbb{E}|x_i(\varphi) - v^{(i-1)}(\varphi)|^2 \leqslant a^2 \sup_{t\in(0,a)} |\varphi^{(i-1)}(t)|^2\sigma_1^2\gamma_1. \tag{52}$$

When $\sigma_1 = 0$, it follows from (52) that

$$\lim_{r\to\infty} \mathbb{E}|x_i(\varphi) - v^{(i-1)}(\varphi)|^2 = 0. \tag{53}$$

Similar to (44)–(48), it can also be obtained that

$$\lim_{r\to\infty} |x_i(\varphi) - v^{(i-1)}(\varphi)| = 0 \text{ almost surely}. \tag{54}$$

This completes the proof of the Theorem 2.

5 Numerical simulations

In this section, the details of numerical simulations are presented to validate the effectiveness of the primary results. Let the input signal be $v(t) = \sin(3t + 1)$. We design a second-order linear TD and a second-order nonlinear TD in the form of (4) in the presence of multiple stochastic disturbances. The linear TD is developed by a linear function given as

$$f(z_1, z_2) = -2z_1 - 4z_2, \ \forall (z_1, z_2) \in \mathbb{R}^2. \tag{55}$$

It can be easily verified that the linear function satisfies Assumption 1 because the matrix $\left[\begin{smallmatrix} 0 & 1 \\ -2 & -4 \end{smallmatrix}\right]$ is Hurwitz. Motivated by the nonlinear feedback controller design, a nonlinear function used in constructing nonlinear TD can be the linear function (55) adding up a Lipschitz continuous function given as

$$f(z_1, z_2) = -2z_1 - 4z_2 - \phi(z_1), \ \forall (z_1, z_2) \in \mathbb{R}^2, \tag{56}$$

where

$$\phi(s) = \begin{cases} -\dfrac{1}{4\pi}, & s \in (-\infty, -\frac{\pi}{2}), \\ \dfrac{1}{4\pi} \sin s, & s \in [-\frac{\pi}{2}, \frac{\pi}{2}], \\ \dfrac{1}{4\pi}, & s \in (\frac{\pi}{2}, +\infty), \end{cases} \tag{57}$$

and Assumption 1 holds for ([22, p.196])

$$\begin{aligned} &V(z_1, z_2) = 1.375z_1^2 + 0.1875z_2^2 + 0.5z_1z_2, \\ &W(z_1, z_2) = 0.5z_1^2 + 0.5z_2^2, \ \lambda_1 = 0.13, \ \lambda_2 = 1.43, \\ &\lambda_3 = \lambda_4 = 0.5, \ c_1 = 3.91, \ c_2 = 2.75. \end{aligned} \tag{58}$$

In Figures 1–3, the relative parameters are specified as

$$\alpha_1 = \alpha_2 = 3, \ \beta_1 = \beta_2 = \frac{1}{18}, \ \sigma_1 = 0.2, \ \sigma_2 = \sigma_3 = 2, \tag{59}$$

and the initial values are taken as

$$x_1(0) = \sin(1), \ x_2(0) = 0, \ w_1(0) = 1, \ w_2(0) = -1. \tag{60}$$

The sampling period $\Delta t = 0.001$, and the diffusion terms $\mathrm{d}B_i(t)$ $(i = 1, 2, 3)$ are simulated by $\sqrt{\Delta t}$ multiplied by random sequences produced by the Matlab program command "randn". The selection of r is as specified by (17). In Figures 1–3, we choose $r = 30$ and $r = 15$, respectively. Further, it can be verified for the nonlinear case that if $r = 15$, $\frac{1}{r} + \frac{1}{2r^{2n-1}} = \frac{1}{15} + \frac{1}{2\times 15^3} \approx 0.07 < \frac{\theta\lambda_3}{\lambda_2} \approx 0.17$, i.e., $r = 15 \in R_0$ and then $r = 30 \in R_0$, where we set $\theta = 0.5$.

It is seen from Figure 1 that the states $x_1(t)$ and $x_2(t)$ of the linear TD track the input signal $v(t) = \sin(3t + 1)$ and the derivative of the input signal satisfactorily almost from the initial stage, respectively. It is also observed from Figure 2 that the tracking effect of the nonlinear TD is at least as good as the linear TD. The fast tracking effect and high tracking accuracy are consistent with the theoretical result that the tracking error becomes small after any given time $T > 0$ with the choice of an appropriate tuning parameter r. Since the tuning parameter r is reduced to be $r = 15$ in Figure 3, it can be seen that the tracking accuracy of the nonlinear TD is relatively not as good as that in Figure 2, which is consistent with the theoretical result that the upper bound of the tracking error is inversely proportional to the tuning parameter r. This observation can also be analyzed quantitatively. It is known that the degree of similarity between two curves can be well characterized by their Hausdorff distance [37]. Indeed, it is calculated by Matlab that the Hausdorff distance between $x_1(t)$ and $\sin(3t + 1)$ is 0.1985 and 0.2673 in Figures 2 and 3, respectively. Furthermore, the Hausdorff distance between $x_2(t)$ and $3\cos(3t + 1)$ is 0.2345 and 0.2658 in Figures 2 and 3, respectively. From the good tracking accuracy in Figures 1 and 2 that two curves are very close to each other. The process noises "$\sigma_2 w_2(t) + \sigma_3 \dot{B}_3(t)$" with large intensity constants $\sigma_2 = \sigma_3 = 2$ do not affect the tracking performance because they are filtered completely, and the input noise cannot be magnified in the tracking process to some extent, which is consistent with the theoretical result. Furthermore, it is observed that the peaking value phenomenon does not exist in Figures 1–3. Finally, in Figures 1–3, the tracking performance of $x_2(t)$ to $\dot{v}(t)$ is not as good as $x_1(t)$ to $v(t)$ because theoretically, the convergence of the former depends on the latter.

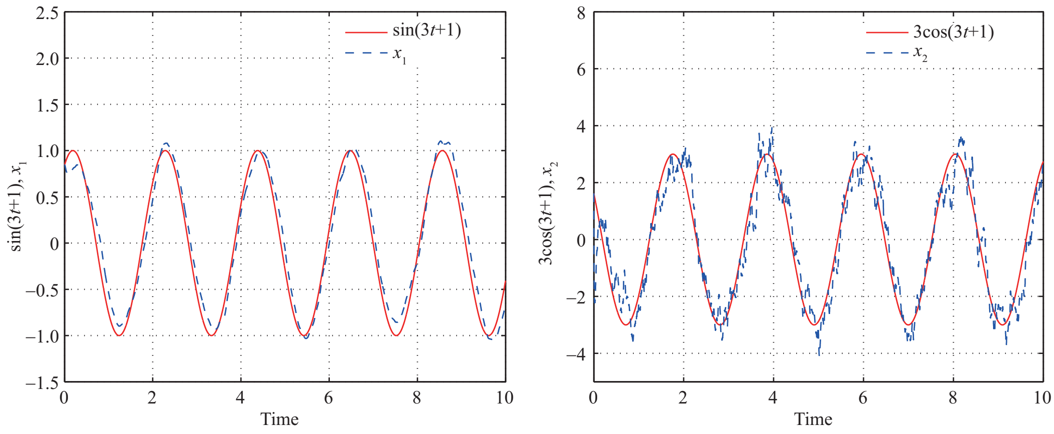

Figure 1 (Color online) The tracking effect of linear TD with $r = 30$.

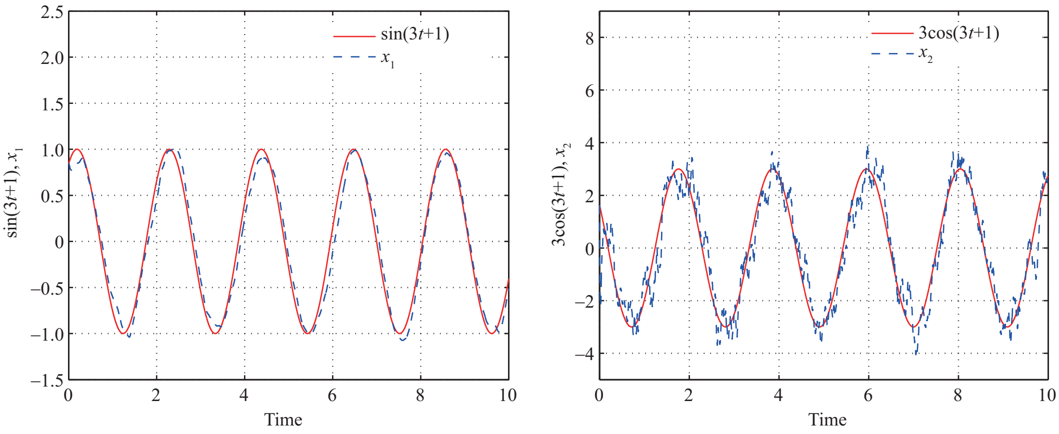

Figure 2 (Color online) The tracking effect of nonlinear TD with $r = 30$.

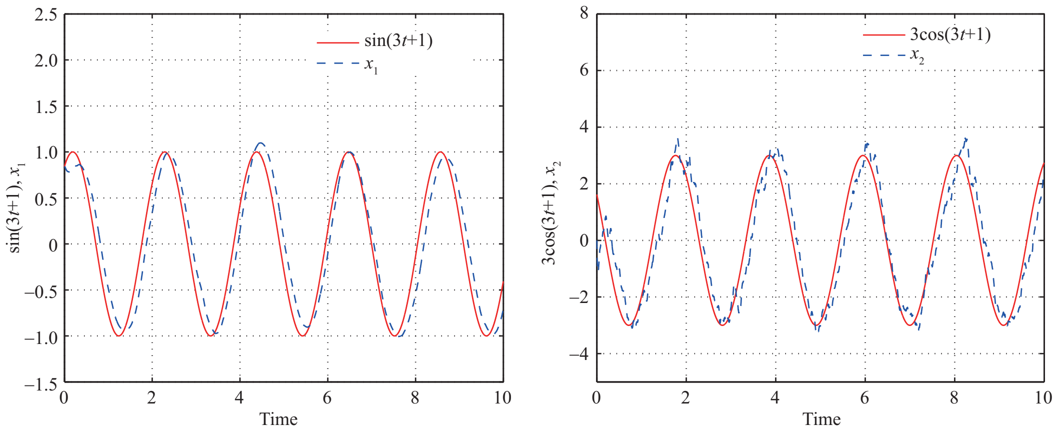

Figure 3 (Color online) The tracking effect of nonlinear TD with $r = 15$.

6 Concluding remarks

The convergence, noise-tolerance, and filtering performance of a TD were investigated, where a general case was considered with the input signal and the TD being disturbed by additive colored noise and additive colored and white noises, respectively. The mathematical proofs are presented to show that the tracking errors of the states of TD to the input signal and its generalized derivatives are convergent in the mean square and even in almost sure sense for the special vanishing input noise. In addition, the details of the numerical simulations are presented to demonstrate the validity of the proposed TD.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61903087, 62173348, 12161141013, 12131008, 62073144, 62333006) and Science and Technology Innovation Program of Hunan Province (Grant No. 2022RC1188).

References

1 Zhao C, Guo L. PID controller design for second order nonlinear uncertain systems. Sci China Inf Sci, 2017, 60: 022201
2 Levant A. Robust exact differentiation via sliding mode technique. Automatica, 1998, 34: 379–384
3 Ibrir S. Linear time-derivative trackers. Automatica, 2004, 40: 397–405
4 Wang X H, Chen Z Q, Yang G. Finite-time-convergent differentiator based on singular perturbation technique. IEEE Trans Automat Contr, 2007, 52: 1731–1737
5 Cruz-Zavala E, Moreno J A, Fridman L M. Uniform robust exact differentiator. IEEE Trans Automat Contr, 2011, 56: 2727–2733
6 Ghanes M, Barbot J P, Fridman L, et al. A new varying-gain-exponent-based differentiator/observer: an efficient balance between linear and sliding-mode algorithms. IEEE Trans Automat Contr, 2020, 65: 5407–5414
7 Rudiger-Wetzlinger M, Reichhartinger M, Horn M. Robust-exact-differentiator-inspired discrete-time differentiation. IEEE Trans Automat Contr, 2022, 67: 3059–3066
8 Han J Q. From PID to active disturbance rejection control. IEEE Trans Ind Electron, 2009, 56: 900–906
9 Han J Q, Wang W. Nonlinear tracking-differentiator. J Syst Sci Mathemat Sci, 1994, 14: 177–183
10 Wang J, Zhang H H, Xiao G X, et al. A comparison study of tracking differentiator and robust exact differentiator. In: Proceedings of the Chinese Automation Congress (CAC), 2020. 1359–1364
11 Xue W C, Huang Y, Yang X X. What kinds of system can be used as tracking-differentiator. In: Proceedings of the 29th Chinese Control Conference, 2010. 6113–6120
12 Zhang H H, Xiao G X, Yu X H, et al. On convergence performance of discrete-time optimal control based tracking differentiator. IEEE Trans Ind Electron, 2021, 68: 3359–3369
13 Dabroom A M, Khalil H K. Output feedback sampled-data control of nonlinear systems using high-gain observers. IEEE Trans Automat Contr, 2001, 46: 1712–1725
14 Kim H W, Sul S K. A new motor speed estimator using Kalman filter in low-speed range. IEEE Trans Ind Electron, 1996, 43: 498–504
15 Su Y X, Zheng C H, Sun D, et al. A simple nonlinear velocity estimator for high-performance motion control. IEEE Trans Ind Electron, 2005, 52: 1161–1169
16 Tian D P, Shen H H, Dai M. Improving the rapidity of nonlinear tracking differentiator via feedforward. IEEE Trans Ind Electron, 2013, 61: 3736–3743
17 Shen J, Xin B, Cui H Q, et al. Control of single-axis rotation INS by tracking differentiator based fuzzy PID. IEEE Trans Aerospace Electron Syst, 2017, 53: 2976–2986
18 Zhang H H, Xie Y D, Xiao G X, et al. A simple discrete-time tracking differentiator and its application to speed and position detection system for a maglev train. IEEE Trans Contr Syst Technol, 2018, 27: 1728–1734
19 Guo B Z, Han J Q, Xi F B. Linear tracking-differentiator and application to online estimation of the frequency of a sinusoidal signal with random noise perturbation. Int J Syst Sci, 2002, 33: 351–358
20 Guo B Z, Zhao Z L. On convergence of tracking differentiator. Int J Control, 2011, 84: 693–701
21 Guo B Z, Zhao Z L. Weak convergence of nonlinear high-gain tracking differentiator. IEEE Trans Automat Contr, 2012, 58: 1074–1080
22 Guo B Z, Zhao Z L. Active Disturbance Rejection Control for Nonlinear Systems: An Introduction. New York: John Wiley & Sons, 2016
23 Deng H, Krstić M. Stochastic nonlinear stabilization-I: a backstepping design. Syst Control Lett, 1997, 32: 143–150
24 Pan Z, Basar T. Backstepping controller design for nonlinear stochastic systems under a risk-sensitive cost criterion. SIAM J Control Optim, 1999, 37: 957–995
25 Deng H, Krstic M. Output-feedback stochastic nonlinear stabilization. IEEE Trans Automat Contr, 1999, 44: 328–333
26 Zhao X Y, Deng F Q, Gao W H. Exponential stability of stochastic Markovian jump systems with time-varying and distributed delays. Sci China Inf Sci, 2021, 64: 209202
27 Cong X R, Zhao C. PID control of uncertain nonlinear stochastic systems with state observer. Sci China Inf Sci, 2021, 64: 192201
28 Lv C W, Ouyang Z Y, Wu Z H, et al. Practical tracking of MIMO uncertain stochastic systems driven by colored noises via active disturbance rejection control. Sci China Inf Sci, 2022, 65: 162208
29 Duan J Q. An Introduction to Stochastic Dynamics (Vol. 51). Cambridge: Cambridge University Press, 2015
30 Kłosek-Dygas M M, Matkowsky B J, Schuss Z. Colored noise in dynamical systems. SIAM J Appl Math, 1988, 48: 425–441
31 Hänggi P, Jung P. Colored noise in dynamical systems. Adv Chem Phys, 1995, 89: 239–326
32 Mao X R. Stochastic Differential Equations and Applications. Chichester: Horwood Publishing Limited, 2007
33 Oksendal B. Stochastic Differential Equations: An Introduction with Applications. Berlin: Springer, 2013
34 Texas Inrumentsts. Technical Reference Manual. TMS320F28069M, TMS320F28068M InstaSPIN$^{\mathrm{TM}}$-MOTION Software, Literature Number: SPRUHJ0A. April 2013, Revised November 2013, 2013
35 Tsinias J. The concept of "exponential input to state stability" for stochastic systems and applications to feedback stabilization. Syst Control Lett, 1999, 36: 221–229
36 Liu S J, Zhang J F, Jiang Z P. A notion of stochastic input-to-state stability and its application to stability of cascaded stochastic nonlinear systems. Acta Math Appl Sin Engl Ser, 2008, 24: 141–156
37 Alt H, Braß P, Godau M, et al. Computing the Hausdorff distance of geometric patterns and shapes. In: Discrete and Computational Geometry. Berlin: Springer, 2003. 65–76