Chaos, Solitons and Fractals 202 (2026) 117597

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Affect of state feedback on tracking performance of two types of nonlinear ILC systems

Yamiao Zhang [a], Jian Liu [b,*], Yuanshi Zheng [b], Xiaoe Ruan [c]

[a] School of Science, Xi'an University of Post and Telecommunications, Xi'an, 710061, China
[b] School of Mechano-electronic Engineering, Xidian University, Xi'an, 710071, China
[c] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China

ARTICLE INFO

Keywords:
Iterative learning control
Tracking performance
Continuous nonlinearity
Local Lipschitz nonlinearity
Nonlinear system

ABSTRACT

This paper considers how state feedback shapes the tracking performance of continuously nonlinear and locally Lipschitz nonlinear ILC systems with measurement noises, respectively. Under spectral radius condition, we first show that the continuous nonlinearity can ensure the bounded tracking performance. Then, under local Lipschitz nonlinearity, we establish the relations between steady-state/transient tracking performance and state feedback as well as the standard deviations of measurement noises, which reveal how state feedback and the standard deviations shape steady-state/transient tracking performance, and how to utilize state feedback to enhance steady-state/transient tracking performance. Finally, we use a numerical example with local Lipschitz nonlinearity to show the theoretical findings.

1. Introduction

In practice, there are many repetitive control systems operated over a fixed finite-time interval such as biomedical engineering [1], batch processes [2], robotics [3] and the references therein, whose task is to achieve precise tracking of some given trajectory point-by-point over the given finite-time interval. The above point-by-point trajectory tracking task is different from the trajectory tracking task considered in the literature [4–6]. Specifically, the systems considered in [4–6] operate over infinite-time interval, and the trajectory tracking task is to track the given reference as time goes to infinity. Iterative learning control, short for ILC, is tailored for the point-by-point trajectory tracking task of repetitive control system over finite-time interval, whose initial idea is to utilize the system input and output data to update the input for better tracking performance [3]. One of the advantages of ILC is that less system dynamics information is required. In view of simple structure and excellent performance of ILC, scholars have conducted extensive and in-depth research in both basic theory [7–11] and practical applications, such as high-speed train systems [12], multi-agent systems [13,14], wind turbine rotor systems [15] and more application scenarios can be found in the literature [7,8,16], to name just a few.

In ILC, the key problem is to construct an update algorithm of the input. Without considering the initial state shift and the load and measurement noises, scholars mainly focus on the tracking performance of ILC algorithm, including transient tracking performance and convergence speed [17–20], to name just a few. In the early works, more attention was paid to designing various ILC algorithms for faster convergence speed [17,20]. Notice that the systems considered in ILC perform repetitive task, and even for discrete-time systems with locally Lipschitz nonlinearity, the simple P-type ILC algorithm can guarantee an exponential

* Corresponding author.
E-mail addresses: yamiao_zhang@163.com (Y. Zhang), ljzmy@xidian.edu.cn (J. Liu), zhengyuanshi2005@163.com (Y. Zheng), wruanxe@mail.xjtu.edu.cn (X. Ruan).

https://doi.org/10.1016/j.chaos.2025.117597
Received 10 August 2025; Received in revised form 20 October 2025; Accepted 10 November 2025
Available online 13 November 2025
0960-0779/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

convergence rate [18]. However, as reported by Lee and Bien in [21], even for linear systems, P-type ILC algorithm cannot prevent the occurrence of huge overshoot. Therefore, to a certain sense, the convergence rate becomes less important than the transient tracking performance. Since 1997, many scholars focus on the design of ILC algorithm for better transient performance so as to avoid the overshoot of the iterative learning process [22–25]. Clearly, for an ILC algorithm, the initial input and learning gains are the key tunable factors that affect transient tracking performance. Generally speaking, the selection of initial input depends on personal experience, and is somewhat random. Using the idea to gradually reduce the magnitude of input updates over time, for continuous-time system, [21] proposed an ILC algorithm with exponentially decaying gain, whose effectiveness was validated by numerical simulations. For discrete-time system, [23] developed an ILC algorithm where the gain is exponentially decaying, and demonstrated that it was able to guarantee the monotone convergence. In practice, the major drawback of the exponential decay gain technique is that it is unable to effectively handle ILC tasks with long trial lengths. [24,25] proposed D-type update law with the piecewise learning mechanism, whose feasibility needs further investigation in the presence of measurement noises.

Notice that the dynamic characteristics of control system along time axis are the key cause of poor transient tracking performance [19] and state feedback can significantly improve the dynamic performance of the system along time-axis. In the case the system state is measurable, incorporating state feedback into ILC update law may be a good choice. Many scholars have done a great deal of valuable work in this area. Using 2-D system theory, [22] developed a synthesis ILC update law for discrete-time linear system, and illustrated that linear state feedback was able to guarantee the output to be the desired trajectory after one iteration. For continuous-time linear system, [26] demonstrated that by introducing linear state feedback into P-type ILC algorithm, it was theoretically possible to ensure that the output is the desired trajectory after one iteration. Both the one-step learning schemes proposed in [22,26] require precise system dynamics. [27] further explored the mechanism of the effect of state feedback on monotone tracking performance of linear learning system. By incorporating nonlinear state feedback into P-type ILC algorithm, [18] proposed a synthesis ILC algorithm for affine-nonlinear SISO systems with continuous nonlinearity and locally Lipschitz nonlinearity, respectively. Theoretical analysis and numerical simulations have demonstrated that state feedback can effectively enhance the transient performance of nonlinear learning system and even eliminate overshoot phenomena. In [28], for affine nonlinear MIMO system, the equivalence and convergence of two ILC algorithms were considered. As illustrated in [28], the two nonlinear ILC algorithms have different application scenarios: one is suitable for ILC system where the trial lengths are iteratively varying and the other is suitable for networked ILC system with packet loss or delay [29,30], to name just a few. The advantages of ILC algorithm comprising state feedback lie in two aspects: a) unlike exponentially decaying gain technique and piecewise learning mechanism, it can take into account both convergence speed and transient tracking performance; b) with enough system dynamics information, it completely cancels the effectiveness of trial lengths on transient tracking performance. It is noted that P-type ILC algorithm comprising state feedback has excellent tracking performance and simple structure and the measurement noises are inevitable in practice. This paper investigates how state feedback shapes tracking performances of P-type ILC algorithm comprising nonlinear state feedback for discrete-time affine nonlinear system with measurement noises, and how to utilize state feedback to enhance the tracking performance. Different from the case without measurement noises, we mainly focus on the relations between nonlinear state feedback and the transient and steady-state tracking performance of learning system.

In summary, this paper contributes in the following three aspects.

(1) By point-wise analysis technique, we first demonstrate that continuous nonlinearity can ensure bounded convergence of P-type ILC system. This surely extends the results presented in [31], which require locally Lipschitz nonlinearity with additional growth constraints for bounded convergence of ILC system developed.

(2) Under local Lipschitz nonlinearity, we establish the relations between steady-state tracking performance and state feedback as well as the standard deviations of measurement noises. The relations reveal how state feedback as well as the standard deviations shapes steady-state tracking performance, and how to utilize state feedback to enhance the steady-state tracking performance.

(3) Under local Lipschitz nonlinearity, we establish the relations between state feedback and the transient tracking performance as well as the standard deviations of measurement noises, which reveal how to utilize state feedback to enhance the transient tracking performance.

The main difference between this article and the existing literature [28] lies in the different issues being studied. Under the condition there is no measurement noise, the authors in [28] investigated the mechanism of the effect of state feedback on the convergence speed and the transient tracking performance of learning system. In this paper, we investigate the mechanism of the effect of state feedback and the standard deviations of measurement noises on the steady-state and transient tracking performance of learning system.

Below is the section arrangement of this paper. Section 2 provides problem description. Section 3 gives the convergence results for the case that the output dimension is equal to that of the input. Section 4 presents the convergence results for the case that the output dimension is less than that of the input. An example with locally Lipschitz nonlinearity is given in Section 5. We conclude this paper in Section 6.

Notations: We denote the set of positive integers by $\mathbb{N}$, the set of n-dimensional real column vectors by $\mathbb{R}^n$ and the set of $m\times n$ real matrices by $\mathbb{R}^{m\times n}$. $T\in\mathbb{N}$ labels the trial length of the considered repetitive control system, $\mathbb{S}=\{0,1,\cdots,T\}$, $\mathbb{S}^+=\mathbb{S}-\{0\}$ and $\mathbb{S}^-=\mathbb{S}-\{T\}$. For $x=[x_1,\ldots,x_n]^\top\in\mathbb{R}^n$, $\|x\|=\sum_{i=1}^n|x_i|$. For $\mathbf{A}=[a_{i,j}]\in\mathbb{R}^{m\times n}$, rank($\mathbf{A}$) refers to its rank, $\mathbf{A}^\top$ labels its transpose and $\|\mathbf{A}\|=\max_{1\le j\le m}\sum_{i=1}^n|a_{i,j}|$. The notation $0_{m\times n}$ denotes the $m\times n$ zero matrix. In addition, for $\mathbf{D}\in\mathbb{R}^{n\times n}$, $\rho(\mathbf{D})$ denotes its spectral radius and $\mathbb{E}\{\cdot\}$ denotes the expectation of random variable '$\cdot$'. 'IOCM' is short for 'input–output coupling matrix' and 'i.i.d.' is short for 'independent and identically distributed'. For nonnegative real numbers $\tilde{a}$ and $\bar{a}$, $\tilde{a}\ll\bar{a}$ means that $\tilde{a}$ is much smaller than $\bar{a}$.

2. Problem description

Consider the nonlinear repetitive system:

$$\begin{cases} x_k(t+1) = \mathbf{f}(t, x_k(t)) + \mathbf{B}u_k(t), t \in \mathbb{S}^-, \\ y_k(t) = \mathbf{C}x_k(t), t \in \mathbb{S}^+, \\ x_k(0) = x^0, \end{cases} \tag{1}$$

where t is the time and k labels the trial number whose range is $\mathbb{N}$. At the kth trial, $x_k(t) \in \mathbb{R}^n$ is the measurable state, $u_k(t) \in \mathbb{R}^p$ is the input and $y_k(t) \in \mathbb{R}^q$ is the output. $\mathbf{B}$ and $\mathbf{C}$ are the input and output matrices with appropriate dimensions, and $\mathbf{f}$ is a nonlinear function from $\mathbb{S}^- \times \mathbb{R}^n$ to $\mathbb{R}^n$. x^0 is a fixed initial state, and $x_k(0) = x^0(\forall k \in \mathbb{N})$ means that each trial has the same initial state x^0.

The definition of the realizable desired trajectory given below is useful [28].

Definition 1. $y_d(t)(t \in \mathbb{S}^+)$ of system (1) is termed a realizable desired trajectory if there is a pair of desired input and state $u_d(t)(t \in \mathbb{S}^-)$ and $x_d(t)(t \in \mathbb{S})$ such that

$$\begin{cases} x_d(t+1) = \mathbf{f}(t, x_d(t)) + \mathbf{B}u_d(t), t \in \mathbb{S}^-, \\ y_d(t) = \mathbf{C}x_d(t), t \in \mathbb{S}^+, \\ x_d(0) = x^0. \end{cases} \tag{2}$$

In the rest of this paper, for the sake of simplicity, we simply denote $u_d(t)(t \in \mathbb{S}^-)$, $y_d(t)(t \in \mathbb{S}^+)$ and $x_d(t)(t \in \mathbb{S})$ as $u_d(\cdot)$, $y_d(\cdot)$ and $x_d(\cdot)$, respectively. We use $\langle u_d(\cdot), x_d(\cdot)\rangle$ to denote the desired input and state pair $u_d(\cdot)$ and $x_d(\cdot)$.

From [27,28], we can get the following properties of desired trajectory realizability of system (1) that will be utilized to establish the convergence results.

P1: Consider system (1), where the IOCM $\mathbf{CB}$ is full-row rank and the dimension of input is equal to that of output. Then, for any $y_d(\cdot)$, there is a unique pair $\langle u_d(\cdot), x_d(\cdot)\rangle$ that satisfies (2), where $u_d(\cdot)$ and $x_d(\cdot)$ can be given by

$$\begin{cases} u_d(t) = [\mathbf{CB}]^{-1}[y_d(t+1) - \mathbf{Cf}(t, x_d(t))], t \in \mathbb{S}^-, \\ x_d(t+1) = \mathbf{f}(t, x_d(t)) + \mathbf{B}u_d(t), t \in \mathbb{S}^-, \\ x_d(0) = x^0. \end{cases} \tag{3}$$

P2: Consider system (1), where $\mathbf{CB}$ is strictly full-row rank, i.e., $\mathbf{CB}$ is full-row rank and the dimension of input is greater than that of output. Then, for any $y_d(\cdot)$, there are infinitely many pairs $\langle u_d(\cdot), x_d(\cdot)\rangle$ that satisfy (2).

P3: Consider system (1), where $\mathbf{CB}$ is strictly full-column rank, i.e., $\mathbf{CB}$ is full-column rank and the dimension of input is less than that of output. Then, the probability of the event that a given desired trajectory of system (1) is realizable is 0.

Remark 1. Apparently, in the case rank($\mathbf{CB}$) $= q = p$, as long as system (1) contains unknown dynamic information, the unique desired input for the given desired trajectory, generated by (3), is not available.

Remark 2. In the case rank($\mathbf{CB}$) $= q < p$, for any given $y_d(\cdot)$, if the sequence $u_k(\cdot)$, short for $u_k(t)(t \in \mathbb{S}^-)$, switches between different desired inputs along with the trial number k, then the control system can precisely track $y_d(\cdot)$. Consequently, even if the sequence $u_k(\cdot)$ can drive system (1) to track a given $y_d(\cdot)$, we cannot assert whether the sequence $u_k(\cdot)$ is convergent when k goes to infinity. The above fact implies that (2) cannot be directly utilized to analyze the convergence properties of ILC system when $\mathbf{CB}$ is strictly full-row rank.

According to the properties presented in **P1**, **P2** and **P3**, if system (1) is expected to have the potential to precisely track any given desired trajectory, the following assumption on $\mathbf{CB}$ is necessary.

Assumption 1. rank($\mathbf{CB}$)=q.

Remark 3. In the existing literature, some authors have studied the 'realizability' of the desired trajectory from different perspectives. For LTI MIMO discrete-time system, using multi-agent theory, the authors in [32] studied the realizability of the desired trajectory from the perspective of the solvability of linear equations. For LTI MIMO continuous-time, by using frequency-domain analysis techniques, the authors in [33] studied the realizability of the desired trajectory from the perspective of trajectory tracking. This paper employs recursive iterative techniques given in [27] to investigate the realizability of the desired trajectory. Although the methods employed in [27,32,33] are not exactly the same, their essence is essentially the same. That is, for the given desired trajectory, the corresponding desired input can be found through the adopted methods.

As shown in the literature (see, e.g., [28]), for system (1), the following ILC algorithm

$$u_{k+1}(t) = u_k(t) + \left[\boldsymbol{\Phi}(t, x_{k+1}(t)) - \boldsymbol{\Phi}(t, x_k(t))\right] + \Gamma[y_d(t+1) - y_k(t+1)] \text{ for } t \in \mathbb{S}^- \tag{4}$$

can fully utilize the known system dynamics information, and improve the transient tracking performance and convergence speed of learning system (1)–(4) simultaneously, where Γ denotes the gain matrix and $\boldsymbol{\Phi}(\cdot,\cdot)$ denotes the nonlinear state feedback item.

In view of the excellent performance of (4) and given that the measurement noises are inevitable in practice, this paper is devoted to investigating the robust convergence of learning system (1)–(5) with measurement noises.

$$u_{k+1}(t) = u_k(t) + [\Phi(t, \hat{x}_{k+1}(t)) - \Phi(t, \hat{x}_k(t))] + \Gamma[y_d(t+1) - \hat{y}_k(t+1)] \text{ for } t \in \mathbb{S}^-, \tag{5}$$

where $\hat{y}_k(t) = y_k(t) + v_k(t)$ is the measurement value of $y_k(t)$ and $v_k(t) = [v_k^1(t), \dots, v_k^q(t)]^\top \in \mathbb{R}^q$ is the output measurement noise vector; $\hat{x}_k(t) = x_k(t) + w_k(t)$ is the measurement value of $x_k(t)$ and $w_k(t) = [w_k^1(t), \dots, w_k^n(t)]^\top \in \mathbb{R}^n$ is the state measurement noise vector.

The following mild assumptions on the measurement noises $v_k(t)$ and $w_k(t)$ will be utilized to establish the main results.

Assumption 2. The measurement noises of output and state are uniformly bounded, i.e.,

$$\begin{cases} |v_k^i(t)| \le C_v \text{ for } t \in \mathbb{S}^+, k \in \mathbb{N} \text{ and } 1 \le i \le q, \\ |w_k^j(t)| \le C_w \text{ for } t \in \mathbb{S}^-, k \in \mathbb{N} \text{ and } 1 \le j \le n. \end{cases} \tag{6}$$

Here, the positive real numbers C_v and C_w are called the measurement accuracies of output and state, respectively.

Assumption 3. For all $t \in \mathbb{S}^-$, $1 \le i \le q$ and $1 \le j \le n$, the means of random variables $v_k^i(t+1)$ and $w_k^j(t)$ are both 0, and their variances are $\sigma_{i,v}^2$ and $\sigma_{j,w}^2$, respectively.

Remark 4. Assumption 1 not only ensures the realizability of the desired trajectory, but also guarantees the solvability of the spectral radius condition $\rho(I_q - CB\Gamma) < 1$. According to the convergence results presented in Lemma 2, Assumption 1 is a necessary condition for ILC system (1)–(4) to be able to perform iterative learning. In ILC, when performing the convergence analysis in the presence of measurement noises, some priori hypotheses on the measurement noises are essential (see, e.g., [31,34–36]). The boundedness assumption on measurement noises can be found in [31,36]. Furthermore, from a practical perspective, a limited input of energy will not result in an infinite output of energy. Considering the bounded nature of the sensor range, the assumption that the measurement noise is bounded is reasonable. Furthermore, in the presence of measurement noises, when analyzing the statistical characteristics of the signals of systems like ILC system (1)–(5), it is a common practice to assume that the measurement noise follows a certain distribution. Under the condition that the measurement noises are i.i.d. and with zero mean and moment of order $2+\delta$, the ILC algorithm proposed in [34] ensures the input sequence to track the desired input almost surely, where δ is a positive real number. In [35], it was assumed that the measurement noises are i.i.d. along the k axis and $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n [v_k^i(t)]^2$ are bounded almost surely for all t and i, then the almost sure convergence of ILC process was derived. Clearly, the model hypothesis of measurement noise, the convergence analysis method and the convergence result are closely related. Compared with the existing literature in ILC, the assumptions 2 and 3 on measurement noises are mild.

The following definitions for locally Lipschitz continuous function and continuous function are useful to establish the main results.

Definition 2. $\mathbf{f} : \mathbb{S}^- \times \mathbb{R}^n \to \mathbb{R}^n$ is called a local Lipschitz continuous function if for all positive real number K and $t \in \mathbb{S}^-$, there is a nonnegative real number $L(t, K)$ termed as the local Lipschitz constant such that for all $\|x\| \le K$ and $\|y\| \le K$

$$\|\mathbf{f}(t, x) - \mathbf{f}(t, y)\| \le L(t, K)\|x - y\|.$$

$\mathbf{f}$ is called a continuous function if for all $t \in \mathbb{S}^-$ and $x \in \mathbb{R}^n$, $\lim_{y\to x} \mathbf{f}(t, y) = \mathbf{f}(t, x)$.

Remark 5. Clearly, any local Lipschitz continuous function $\mathbf{f}$ has infinitely many local Lipschitz constants for any t and K. Note that when t and K are fixed,

$$\tilde{L}(t, K) = \sup_{x \ne y, \|x\| \le K, \|y\| \le K} \frac{\|\mathbf{f}(t, x) - \mathbf{f}(t, y)\|}{\|x - y\|}$$

is a nonnegative constant such that

$$\|\mathbf{f}(t, x) - \mathbf{f}(t, y)\| \le \tilde{L}(t, K)\|x - y\| \ (\forall x, y \in \{z \in \mathbb{R}^n : \|z\| \le K\}).$$

This means that for all $K > 0$ and $t \in \mathbb{S}^-$, the nonnegative constant $\tilde{L}(t, K)$ defined above is a local Lipschitz constant of $\mathbf{f}$. It is easy to check that for any local Lipschitz constant $L(t, K)$, $\tilde{L}(t, K) \le L(t, K)$. Consequently, $\tilde{L}(t, K)$ is the smallest local Lipschitz constant of $\mathbf{f}$ for fixed t and K. The smallest Lipschitz constant can more accurately reflect the local variation of the local Lipschitz function. In the following, when we talk about the local Lipschitz constant of a function, we mean the smallest one. Undoubtedly, the local Lipschitz nonlinearity considered in Definition 2 is different from the one considered in [31]. We do not require the function $\mathbf{f}$ and its local Lipschitz constant $L(t, K)$ to satisfy additional growth constraints. In other words, the local Lipschitz nonlinearity considered in this paper is more general.

The lemmas given below are useful to establish the main results. The proof of Lemma 1 can be found in appendix and following the point-wise analysis method used in [28] gives Lemma 2.

Lemma 1. *Suppose that e_k is a nonnegative sequence satisfying*

$$e_{k+1} \le \theta e_k + \phi_k \ (\forall k \in \mathbb{N}), \tag{7}$$

where $\theta(\in [0,1))$ is a constant and ϕ_k is a bounded nonnegative sequence. Then, the nonnegative sequence e_k satisfies

$$e_k \le \max\left\{ e_1, \frac{C_\phi}{1-\theta} \right\} \ (\forall k \in \mathbb{N}), \tag{8}$$

where C_ϕ is the supremum of nonnegative sequence ϕ_k.

Lemma 2. *Consider learning system (1)–(4), where Φ and $\mathbf{f}$ are continuous and rank$(\mathbf{CB}) = q$. Then, $\rho(I_q - \mathbf{CB}\Gamma) < 1$ if and only if for any $y_d(\cdot)$ and $u_1(\cdot)$,*

- *$y_k(t)$ converges to $y_d(t)$ over $\mathbb{S}^+$;*
- *the sequences $u_k(\cdot)$ and $x_k(\cdot)$ converge along the k axis whose limits $u_d(\cdot)$ and $x_d(\cdot)$ together with $y_d(\cdot)$ satisfy Eq. (2).*

3. Convergence results for the case $q = p$

This section aims to establish the convergence results of learning system (1)–(5) for the case that **CB** is full-row rank and the dimension of output is equal to that of input, where continuous nonlinearity and local Lipschitz continuous nonlinearity are considered successively.

In the remainder, for brevity, let $F(\sigma_w, \sigma_v) = L_\Phi \sigma_w + \sigma_v$,

$$e_k^x(t) = x_d(t) - x_k(t), \Delta x_k(t) = x_{k+1}(t) - x_k(t),$$

$$e_k^y(t) = y_d(t) - y_k(t), e_k^u(t) = u_d(t) - u_k(t),$$

$$\delta\mathbf{f}(t, x_k(t)) = \mathbf{f}(t, x_d(t)) - \mathbf{f}(t, x_k(t)),$$

$$\Delta\Phi(t, x_k(t)) = \Phi(t, x_{k+1}(t)) - \Phi(t, x_k(t)),$$

$$\Delta_w\Phi(t, x_k(t)) = \Phi(t, \hat{x}_k(t)) - \Phi(t, x_k(t)),$$

$$\mathbf{f}_\Phi^{\mathbf{B}}(t, x_k(t)) = \mathbf{f}(t, x_k(t)) + \mathbf{B}\Phi(t, x_k(t)),$$

$$\mathbf{f}_\Phi^{\mathbf{CB}}(t, x_k(t)) = \mathbf{Cf}(t, x_k(t)) + \mathbf{CB}\Phi(t, x_k(t)),$$

$$\delta\mathbf{f}_\Phi^{\mathbf{B}}(t, x_k(t)) = \mathbf{f}_\Phi^{\mathbf{B}}(t, x_d(t)) - \mathbf{f}_\Phi^{\mathbf{B}}(t, x_k(t)),$$

$$\delta\mathbf{f}_\Phi^{\mathbf{CB}}(t, x_k(t)) = \mathbf{f}_\Phi^{\mathbf{CB}}(t, x_d(t)) - \mathbf{f}_\Phi^{\mathbf{CB}}(t, x_k(t)).$$

Theorem 1 gives the bounded convergence result for the case $\mathbf{f}$ and Φ are continuous.

Theorem 1. *Consider learning system (1)–(5), where*

(a1) $\mathbf{f}$ and Φ are continuous;

(b1) $\rho(I_q - \mathbf{CB}\Gamma) < 1$ and $q = p$;

(c1) Assumption 2 holds.

Then, for any $y_d(\cdot)$, $u_1(\cdot)$ and $t(\in \mathbb{S}^-)$, there are positive real numbers C_t^y, C_t^u and C_t^x independent of the distributions of measurement noises such that

$$\|e_k^y(t+1)\| \le C_t^y, \|e_k^u(t)\| \le C_t^u \text{ and } \|e_k^x(t+1)\| \le C_t^x \ (\forall k), \tag{9}$$

where $\langle u_d(\cdot), x_d(\cdot)\rangle$ is the unique pair generated by (3).

Remark 6. It follows from (b1) that rank$(\mathbf{CB}) = q = p$. According to the property **P1**, for any desired trajectory, there is a unique desired input that is the only learning goal of system (1)–(5). Since the continuous nonlinearity is considered in Theorem 1, the contraction mapping-based convergence analysis method used in [31] cannot be directly used, and we will adapt a point-wise analysis technique to prove Theorem 1.

Proof. Let $y_d(\cdot)$ and $u_1(\cdot)$ be arbitrarily given. Clearly, $y_d(\cdot)$ and $u_1(\cdot)$ are bounded.

Using (b1) yields rank$(\mathbf{CB}) = q = p$. It follows from **P1** that there is a unique pair $\langle u_d(\cdot), x_d(\cdot)\rangle$ that satisfies (2). The boundedness of $u_d(\cdot)$ and $x_d(\cdot)$ is obvious.

Next, we use mathematical induction to complete the proof. In the remainder, for brevity, when the ranges of k and t are clear, we omit it.

Since $\rho(I_q - \mathbf{CB}\Gamma) < 1$, we can find an invertible matrix $\Xi \in \mathbb{R}^{q\times q}$ that satisfies

$$\sigma = \|\Xi[I_q - \mathbf{CB}\Gamma]\Xi^{-1}\| < 1, \tag{10}$$

where Ξ^{-1} denotes the inverse of Ξ.

Using (1) and (5) yields

$$e^y_{k+1}(t+1) = \left[I_q - \mathbf{CB}\Gamma\right]e^y_k(t+1) + \delta\mathbf{f}^{\mathbf{CB}}_{\Phi}(t, x_{k+1}(t)) - \delta\mathbf{f}^{\mathbf{CB}}_{\Phi}(t, x_k(t)) + \mathbf{CB}\Gamma v_k(t+1) - \mathbf{CB}\left[\Delta_w\Phi(t, x_{k+1}(t)) - \Delta_w\Phi(t, x_k(t))\right], \tag{11}$$

which together with a simple manipulation yields that $\forall t \in \mathbb{S}^-$,

$$\|\Xi e^y_{k+1}(t+1)\| \le \sigma\|\Xi e^y_k(t+1)\| + \varphi_k(t), \tag{12}$$

where

$$\varphi_k(t) = \|\Xi\|\|\delta\mathbf{f}^{\mathbf{CB}}_{\Phi}(t, x_{k+1}(t))\| + \|\Xi\|\|\delta\mathbf{f}^{\mathbf{CB}}_{\Phi}(t, x_k(t))\| + \|\Xi\mathbf{CB}\|\left[\|\Delta_w\Phi(t, x_{k+1}(t))\| + \|\Delta_w\Phi(t, x_k(t))\|\right] + \|\Xi\mathbf{CB}\Gamma\|\|v_k(t+1)\|. \tag{13}$$

Using (1) and (2) gives

$$\begin{cases} e^u_k(t) = [\mathbf{CB}]^{-1}\left[e^y_k(t+1) - \mathbf{C}\delta\mathbf{f}(t, x_k(t))\right], \\ e^x_k(t+1) = \delta\mathbf{f}(t, x_k(t)) + \mathbf{B}e^u_k(t), \end{cases}$$

which implies that $\forall t \in \mathbb{S}^-$,

$$\|e^u_k(t)\| \le \|[\mathbf{CB}]^{-1}\|\left[\|e^y_k(t+1)\| + \|\mathbf{C}\|\|\delta\mathbf{f}(t, x_k(t))\|\right] \tag{14}$$

and

$$\|e^x_k(t+1)\| \le \|\delta\mathbf{f}(t, x_k(t))\| + \|\mathbf{B}\|\|e^u_k(t)\|. \tag{15}$$

We first show that we can find positive real numbers C^y_0, C^u_0, and C^x_0 independent of the distributions of measurement noises such that (9) holds for $t = 0$.

By (13) and $x_k(0) = x_d(0) = x^0$ we have

$$\varphi_k(0) = \|\Xi\mathbf{CB}\|\|\Delta_w\Phi(t, x_{k+1}(0))\| + \|\Xi\mathbf{CB}\|\|\Delta_w\Phi(t, x_k(0))\| + \|\Xi\mathbf{CB}\Gamma\|\|v_k(1)\|.$$

Note that

$$\|\Delta_w\Phi(0, x_k(0))\| = \|\Phi(0, x^0 + w_k(0)) - \Phi(0, x^0)\|,$$

$\|w_k(0)\| \le nC_w$, $\|v_k(1)\| \le qC_v$ and $\Phi(0, x)$ is continuous. Therefore, there is a positive constant C^{φ}_0 independent of the distributions of measurement noises such that

$$\varphi_k(0) \le C^{\varphi}_0 \ (\forall k \in \mathbb{N}). \tag{16}$$

It follows from (10), (12), (16) and Lemma 1 that there is a positive constant $\hat{C}^y_0$ independent of the distributions of measurement noises such that

$$\|\Xi e^y_k(1)\| \le \hat{C}^y_0 \ (\forall k \in \mathbb{N}),$$

which together $\|e^y_k(1)\| \le \|\Xi^{-1}\|\|\Xi e^y_k(1)\|$ implies

$$\|e^y_k(1)\| \le C^y_0 \ (\forall k \in \mathbb{N}), \tag{17}$$

where $C^y_0 = \|\Xi^{-1}\|\hat{C}^y_0$ is independent of the distributions of measurement noises.

By (14), (15) and $x_k(0) = x^0(\forall k)$, we get

$$\|e^u_k(0)\| \le \|[\mathbf{CB}]^{-1}\|\|e^y_k(1)\| \text{ and } \|e^x_k(1)\| \le \|\mathbf{B}\|\|e^u_k(0)\|,$$

which together with (17) gives rise to

$$\|e^u_k(0)\| \le C^u_0 \text{ and } \|e^x_k(1)\| \le C^x_0 \ (\forall k), \tag{18}$$

where $C^u_0 = \|[\mathbf{CB}]^{-1}\|C^y_0$ and $C^x_0 = \|\mathbf{B}\|C^u_0$ are independent of the distributions of measurement noises. It follows from (17) and (18) that there are positive constants C^y_0, C^u_0 and C^x_0 independent of the distributions of measurement noises such that (9) holds for $t = 0$.

Suppose that for $t = t_1 (t_1 \in \mathbb{S}^-)$, there are positive real numbers $C^y_{t_1}$, $C^u_{t_1}$ and $C^x_{t_1}$ independent of the distributions of measurement noises such that (9) holds for $t = t_1$. What the rest we need to do is to show that there are positive real numbers $C^y_{t_1+1}$, $C^u_{t_1+1}$ and $C^x_{t_1+1}$ independent of the distributions of measurement noises such that (9) holds for $t = t_1 + 1$.

Note that $\mathbf{f}$ and Φ are continuous, $\|w_k(t_1+1)\| \le nC_w(\forall k)$ and $\|v_k(t_1+2)\| \le qC_v(\forall k)$. According to (13) and the induction hypothesis, there exists a positive real number $C^{\varphi}_{t_1+1}$ independent of the distributions of measurement noises such that

$$\varphi_k(t_1+1) \le C^{\varphi}_{t_1+1} \ (\forall k \in \mathbb{N}). \tag{19}$$

According to (10), (12), (19) and Lemma 1, we can find a positive real number $\hat{C}^y_{t_1+1}$ independent of the distributions of measurement noises such that

$$\|\Xi e^y_k(t_1+2)\| \le \hat{C}^y_{t_1+1} \ (\forall k \in \mathbb{N}),$$

which together with

$$\|e_k^y(t_1+2)\| \le \|\Xi^{-1}\|\|\Xi e_k^y(t_1+2)\|$$

yields

$$\|e_k^y(t_1+2)\| \le C_{t_1+1}^y \ (\forall k \in \mathbb{N}), \tag{20}$$

where $C_{t_1+1}^y = \|\Xi^{-1}\|\hat{C}_{t_1+1}^y$ is independent of the distributions of measurement noises.

In (14) and (15), replacing t with t_1+1 yields

$$\begin{cases} \|e_k^u(t_1+1)\| \le \|[\mathbf{CB}]^{-1}\|\left[\|e_k^y(t_1+2)\| + \|\mathbf{C}\|\|\delta\mathbf{f}(t, x_k(t_1+1))\|\right], \\ \|e_k^x(t_1+2)\| \le \|\delta\mathbf{f}(t, x_k(t_1+1))\| + \|\mathbf{B}\|\|e_k^u(t_1+1)\|, \end{cases}$$

which together with (20), the induction hypothesis and the continuity of $\mathbf{f}$ gives that there exists positive constants $C_{t_1+1}^u$ and $C_{t_1+1}^x$ independent of the distributions of measurement noises such that

$$\|e_k^u(t_1+1)\| \le C_{t_1+1}^u \text{ and } \|e_k^x(t_1+2)\| \le C_{t_1+1}^x \ (\forall k). \tag{21}$$

Consequently, there are positive real numbers $C_{t_1+1}^y$, $C_{t_1+1}^u$ and $C_{t_1+1}^x$ independent of the distributions of measurement noises such that (9) holds for $t = t_1+1$.

By the mathematical induction, we can find positive real numbers C_t^y, C_t^u and C_t^x $(t \in \mathbb{S}^-)$ independent of the distributions of measurement noises such that (9) holds. □

Remark 7. When $\Phi = 0$, the update law (5) becomes

$$u_{k+1}(t) = u_k(t) + \Gamma[y_d(t+1) - \hat{y}_k(t+1)]. \tag{22}$$

Theorem 1 tells that for learning system (1)–(22), the continuity of $\mathbf{f}$ can guarantee the uniformly bounded convergence in the presence of bounded output measurement noises. Moreover, in learning system (1)–(5), when the bounded load disturbances are involved, along the same technique route it is easy to show that the continuity of $\mathbf{f}$ and Φ is also enough to ensure the bounded convergence. This surely extends the result presented in [31]. In [31], the authors derived the bounded convergence on the premise that $\mathbf{f}$ is a locally Lipschitz function with addition growth constraints.

The following theorem shows the influences of state feedback, gain matrix and standard deviations on steady-state tracking performance of learning system (1)–(5) under the condition that $\mathbf{f}$ and Φ are locally Lipschitz continuous.

Theorem 2. *Consider learning system (1)–(5), where*

(a2) $\mathbf{f}$ *and* Φ *are locally Lipschitz continuous;*

(b2) $\rho(I_q - \mathbf{CB}\Gamma) < 1$ *and* $q = p$;

(c2) Assumptions 2 and 3 hold.

Then, for any $y_d(\cdot)$ *and* $u_1(\cdot)$*, we can find positive real numbers* ϕ_t^y*,* ϕ_t^u *and* $\phi_t^x (t \in \mathbb{S}^-)$ *independent of* Φ *and the distribution of measurement noises such that* $\forall t \in \mathbb{S}^-$*,*

$$\begin{cases} \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^y(t+1)\|\} \le \phi_t^y\left[\sum\limits_{s=0}^{t}\left(L_{\mathbf{Cf}}^{\Phi}\right)^s\right]F(\sigma_w, \sigma_v), \\ \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^u(t)\|\} \le \phi_t^u\left[\sum\limits_{s=0}^{t}\left(L_{\mathbf{Cf}}^{\Phi}\right)^s\right]F(\sigma_w, \sigma_v), \\ \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^x(t+1)\|\} \le \phi_t^x\left[\sum\limits_{s=0}^{t}\left(L_{\mathbf{Cf}}^{\Phi}\right)^s\right]F(\sigma_w, \sigma_v), \end{cases} \tag{23}$$

where $\langle u_d(\cdot), x_d(\cdot)\rangle$ *is the unique pair generated by (3),* $\sigma_w = \sum_{i=1}^n \sigma_{i,w}$*,* $\sigma_v = \sum_{i=1}^q \sigma_{i,v}$ *and* $F(\sigma_w, \sigma_v) = L_{\Phi}\sigma_w + \sigma_v$*.*

Proof. Let $y_d(\cdot)$ and $u_1(\cdot)$ be arbitrarily given.

By (b2), we have a unique pair $\langle u_d(\cdot), x_d(\cdot)\rangle$ that satisfies (2). The boundedness of $y_d(\cdot)$, $u_d(\cdot)$, $x_d(\cdot)$ and $u_1(\cdot)$ is obvious.

Clearly, under the conditions $(a2)$, $(b2)$ and $(c2)$, Theorem 1 holds. As done in the proof of Theorem 1, choose the nonsingular matrix $\Xi \in \mathbb{R}^{q\times q}$ satisfying (10). Therefore,

$$\eta = 1 - \|\Xi[I_q - \Gamma\mathbf{CB}]\Xi^{-1}\| \in (0, 1]. \tag{24}$$

According to Theorem 1, we can find a positive real number K independent of the distributions of measurement noises such that $\|x_d(t)\| \le K$, $\|x_k(t)\| \le K$ and $\|\hat{x}_k(t)\| \le K$ $\forall t, k$. Due to $(a2)$, there are nonnegative constants $L_{\mathbf{Cf}}^{\Phi}$, $L_{\mathbf{f}}$, $L_{\mathbf{f}}^{\Phi}$ and L_{Φ} that satisfy

$$\|\mathbf{f}_{\Phi}^{\mathbf{CB}}(t, z_1) - \mathbf{f}_{\Phi}^{\mathbf{CB}}(t, z_2)\| \le L_{\mathbf{Cf}}^{\Phi}\|z_1 - z_2\|, \tag{25}$$

$$\|\mathbf{f}(t, z_1) - \mathbf{f}(t, z_2)\| \le L_{\mathbf{f}}\|z_1 - z_2\|, \tag{26}$$

$$\|\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{B}}(t,z_1)-\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{B}}(t,z_2)\|\le L_{\mathbf{f}}^{\boldsymbol{\Phi}}\|z_1-z_2\|, \tag{27}$$

$$\|\boldsymbol{\Phi}(t,z_1)-\boldsymbol{\Phi}(t,z_1)\|\le L_{\boldsymbol{\Phi}}\|z_1-z_2\| \tag{28}$$

$\forall z_1,z_2\in\{z:\|z\|\le K\}$ and $t\in\mathbb{S}^-$. Noticing that K is independent of the distributions of measurement noises, $L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}$, $L_{\mathbf{f}}$, $L_{\mathbf{f}}^{\boldsymbol{\Phi}}$ and $L_{\boldsymbol{\Phi}}$ are independent of the distributions of measurement noises. By $\mathbb{E}\{|v_k^i(t)|^2\}=\sigma_{i,v}^2$, $\mathbb{E}\{|w_k^i(t)|^2\}=\sigma_{i,w}^2$ and $\sigma_w=\sum_{i=1}^n\sigma_{i,w}$, one has

$$\mathbb{E}\{\|v_k(t+1)\|\}\le\sigma_v \text{ and } \mathbb{E}\{\|w_k(t)\|\}\le\sigma_w \tag{29}$$

$\forall t\in\mathbb{S}^-$ and $k\in\mathbb{N}$.

Using (12), (13), (14) and (15) together with (25), (26) and (28) gives

$$\|\Xi e_{k+1}^y(t+1)\|\le\sigma\|\Xi e_k^y(t+1)\|+\|\Xi\|L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}\left[\|e_{k+1}^x(t)\|+\|e_k^x(t)\|\right]+\|\Xi\mathbf{CB}\|L_{\boldsymbol{\Phi}}\left[\|w_{k+1}(t)\|+\|w_k(t)\|\right]+\|\Xi\mathbf{CB}\Gamma\|\|v_k(t+1)\|,$$

$$\|e_k^u(t)\|\le\|[\mathbf{CB}]^{-1}\|\left[\|e_k^y(t+1)\|+\|\mathbf{C}\|L_{\mathbf{f}}\|e_k^x(t)\|\right]$$

and

$$\|e_k^x(t+1)\|\le L_{\mathbf{f}}\|e_k^x(t)\|+\|\mathbf{B}\|\|e_k^u(t)\|,$$

which together with (29) yield that $\forall t\in\mathbb{S}^-$,

$$\mathbb{E}\{\|\Xi e_{k+1}^y(t+1)\|\}\le 2\|\Xi\mathbf{CB}\|L_{\boldsymbol{\Phi}}\sigma_w+\|\Xi\mathbf{CB}\Gamma\|\sigma_v+\sigma\mathbb{E}\{\|\Xi e_k^y(t+1)\|\}+\|\Xi\|L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}\left[\mathbb{E}\{\|e_{k+1}^x(t)\|\}+\mathbb{E}\{\|e_k^x(t)\|\}\right], \tag{30}$$

$$\mathbb{E}\{\|e_k^u(t)\|\}\le\|[\mathbf{CB}]^{-1}\|\left[\mathbb{E}\{\|e_k^y(t+1)\|\}+\|\mathbf{C}\|L_{\mathbf{f}}\mathbb{E}\{\|e_k^x(t)\|\}\right] \tag{31}$$

and

$$\mathbb{E}\{\|e_k^x(t+1)\|\}\le L_{\mathbf{f}}\mathbb{E}\{\|e_k^x(t)\|\}+\|\mathbf{B}\|\mathbb{E}\{\|e_k^u(t)\|\}. \tag{32}$$

Now, we use mathematical induction w.r.t time t to prove (23).

By (24), (30) and $x_k(0)=x^0(\forall k)$, we have

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|\Xi e_k^y(1)\|\}\le\frac{\|\Xi\mathbf{CB}\|}{\eta}\left[2L_{\boldsymbol{\Phi}}\sigma_w+\|\Gamma\|\sigma_v\right],$$

which together with

$$\mathbb{E}\{\|e_k^y(t)\|\}\le\|\Xi^{-1}\|\mathbb{E}\{\|\Xi e_k^y(t)\|\} \tag{33}$$

gives rise to

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|e_k^y(1)\|\}\le\phi_0^y\left[L_{\boldsymbol{\Phi}}\sigma_w+\sigma_v\right], \tag{34}$$

where $\phi_0^y=\frac{\|\Xi^{-1}\|\|\Xi CB\|}{\eta}\max\{2,\|\Gamma\|\}$ is independent of $\boldsymbol{\Phi}$ and the distribution of measurement noises $w_k(t)$ and $v_k(t)$.

Using (31), (34) and $x_k(0)=x^0(\forall k)$ together with a simple computation yields

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|\delta u_k(0)\|\}\le\phi_0^u\left[L_{\boldsymbol{\Phi}}\sigma_w+\sigma_v\right], \tag{35}$$

where $\phi_0^u=\|[\mathbf{CB}]^{-1}\|\phi_0^y$ are independent of $\boldsymbol{\Phi}$ and the distribution of measurement noises. Then, by (32), (35) and $x_k(0)=x^0(\forall k)$, we have

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|e_k^x(1)\|\}\le\phi_0^x\left[L_{\boldsymbol{\Phi}}\sigma_w+\sigma_v\right], \tag{36}$$

where $\phi_0^x=\|\mathbf{B}\|\phi_0^u$ are independent of $\boldsymbol{\Phi}$ and the distribution of measurement noises. It follows from (34), (35) and (36) that (23) is true for $t=0$.

Suppose that (23) is true for $t=t_1(t_1<T-1)$, i.e., there are nonnegative constants $\phi_{t_1}^y$, $\phi_{t_1}^u$ and $\phi_{t_1}^x$ independent of $\boldsymbol{\Phi}$ and the distribution of measurement noises such that

$$\begin{cases}\lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e_k^y(t_1+1)\|\}\le\phi_{t_1}^y\left[\sum\limits_{s=0}^{t_1}\left(L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}\right)^s\right]F(\sigma_w,\sigma_v),\\ \lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e_k^u(t_1)\|\}\le\phi_{t_1}^u\left[\sum\limits_{s=0}^{t_1}\left(L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}\right)^s\right]F(\sigma_w,\sigma_v),\\ \lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e_k^x(t_1+1)\|\}\le\phi_{t_1}^x\left[\sum\limits_{s=0}^{t_1}\left(L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}\right)^s\right]F(\sigma_w,\sigma_v).\end{cases} \tag{37}$$

Next, we show that there are nonnegative constants $\phi_{t_1+1}^y$, $\phi_{t_1+1}^u$ and $\phi_{t_1+1}^x$ independent of $\boldsymbol{\Phi}$ and the distribution of measurement noises such that (23) holds for $t=t_1+1$.

In (30), (31) and (32), replacing t with t_1+1 gives

$$\mathbb{E}\{\|\Xi e^y_{k+1}(t_1+2)\|\} \le \sigma\mathbb{E}\{\|\Xi e^y_k(t_1+2)\|\} + 2\|\Xi\mathbf{CB}\|L_{\Phi}\sigma_w + \|\Xi\mathbf{CB}\Gamma\|\sigma_v + \|\Xi\|L^{\Phi}_{\mathbf{Cf}}\big[\mathbb{E}\{\|e^x_k(t_1+1)\|\} + \mathbb{E}\{\|e^x_{k+1}(t_1+1)\|\}\big], \tag{38}$$

$$\mathbb{E}\{\|e^u_k(t_1+1)\|\} \le \|[\mathbf{CB}]^{-1}\|\ \big[\ \mathbb{E}\{\|e^y_k(t_1+2)\|\} + \|\mathbf{C}\|L_{\mathbf{f}}\mathbb{E}\{\|e^x_k(t_1+1)\|\}\ \big], \tag{39}$$

and

$$\mathbb{E}\{\|e^x_k(t_1+2)\|\} \le L_{\mathbf{f}}\mathbb{E}\{\|e^x_k(t_1+1)\|\} + \|\mathbf{B}\|\mathbb{E}\{\|e^u_k(t_1+1)\|\}. \tag{40}$$

Using (38) and the induction hypothesis (37), we have

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|\Xi e^y_k(t_1+2)\|\} \le \max\Big\{\frac{2\|\Xi\mathbf{CB}\|}{\eta}, \frac{\|\Xi\mathbf{CB}\Gamma\|}{\eta}, \frac{2\|\Xi\|}{\eta}\phi^x_{t_1}\Big\}\Big[\sum_{s=0}^{t_1+1}\big(L^{\Phi}_{\mathbf{Cf}}\big)^s\Big]F(\sigma_w,\sigma_v),$$

which together with (33) implies

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|e^y_k(t_1+2)\|\} \le \phi^y_{t_1+1}\Big[\sum_{s=0}^{t_1+1}\big(L^{\Phi}_{\mathbf{Cf}}\big)^s\Big]F(\sigma_w,\sigma_v), \tag{41}$$

where

$$\phi^y_{t_1+1} = \|\Xi^{-1}\|\max\Big\{\frac{2\|\Xi\mathbf{CB}\|}{\eta}, \frac{\|\Xi\mathbf{CB}\Gamma\|}{\eta}, \frac{2\|\Xi\|}{\eta}\phi^x_{t_1}\Big\}$$

are independent of Φ and the distribution of measurement noises.

By (39), (37) and (41), we obtain

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|e^u_k(t_1+1)\|\} \le \phi^u_{t_1+1}\Big[\sum_{s=0}^{t_1+1}\big(L^{\Phi}_{\mathbf{Cf}}\big)^s\Big]F(\sigma_w,\sigma_v), \tag{42}$$

where $\phi^u_{t_1+1} = \|[\mathbf{CB}]^{-1}\|\big[\phi^y_{t_1+1} + \|\mathbf{C}\|L_{\mathbf{f}}\phi^x_{t_1}\big]$ is independent of ϕ and the distribution of measurement noises.

It follows from (40) and (42) together with the induction hypothesis (37) that

$$\lim_{k\to\infty}\sup\mathbb{E}\{\|e^x_k(t_1+1)\|\} \le \phi^x_{t_1+1}\Big[\sum_{s=0}^{t_1+1}\Big(L^{\Phi}_{\mathbf{Cf}}\Big)^s\Big]F(\sigma_w,\sigma_v), \tag{43}$$

where $\phi^x_{t_1+1} = L_{\mathbf{f}}\phi^x_{t_1} + \|\mathbf{B}\|\phi^u_{t_1+1}$ are independent of Φ and the distribution of measurement noises.

According to (41), (42) and (43), for $t = t_1+1$, there are nonnegative constants $\phi^y_{t_1+1}$, $\phi^u_{t_1+1}$ and $\phi^x_{t_1+1}$, independent of Φ and the distribution of measurement noises, such that (23) is true for $t = t_1+1$.

By mathematical induction, we have proven that for all $t\in\mathbb{S}^-$, there are nonnegative constants ϕ^y_t, ϕ^u_t and ϕ^x_t, independent of Φ and the measurement noises, such that (23) holds. □

Remark 8. Theorem 2 tells that under the local Lipschitz continuity, the output, input and state steady-state tracking performances of learning system (1)–(5) are governed by bilinear functions of the standard deviations of the state and output measurement noises, respectively. When the output and state measurement accuracies C_v and C_w are fixed, the smaller the standard deviations of the measurement noises are, the better the expected steady-state tracking performances may be. When $\Phi = 0$, (23) becomes

$$\begin{cases}\lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e^y_k(t+1)\|\} \le \phi^y_t\big[\sum_{s=0}^{t}\big(L_{\mathbf{Cf}}\big)^s\big]\sigma_v,\\ \lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e^u_k(t)\|\} \le \phi^u_t\big[\sum_{s=0}^{t}\big(L_{\mathbf{Cf}}\big)^s\big]\sigma_v,\\ \lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e^x_k(t+1)\|\} \le \phi^x_t\big[\sum_{s=0}^{t}\big(L_{\mathbf{Cf}}\big)^s\big]\sigma_v.\end{cases} \tag{44}$$

(44) tells that when $\Phi = 0$ and $L_{\mathbf{Cf}} > 1$, the steady-state tracking performance of ILC system (1)–(5) may deteriorate as T increases.

Remark 9. According to (23), a smaller Lipschitz constant $L^{\Phi}_{\mathbf{Cf}}$ may imply better steady-state tracking performance, where $L^{\Phi}_{\mathbf{Cf}}$ is a local Lipschitz constant of $\mathbf{Cf}+\mathbf{CB}\Phi$. Clearly, $L^{\Phi}_{\mathbf{Cf}}$ is determined by the supremum of system states during the iterative learning process, the function $\mathbf{Cf}$, the IOCM $\mathbf{CB}$, and the state feedback item Φ. In practice, when $\mathbf{Cf}$ and $\mathbf{CB}$ are known, we can directly let $\Phi = -[\mathbf{CB}]^{\top}\big(\mathbf{CB}[\mathbf{CB}]^{\top}\big)^{-1}\mathbf{Cf}$. Then, (23) becomes

$$\begin{cases}\lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e^y_k(t+1)\|\} \le \phi^y_t[L_{\Phi}\sigma_w + \sigma_v],\\ \lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e^u_k(t)\|\} \le \phi^u_t[L_{\Phi}\sigma_w + \sigma_v],\\ \lim\limits_{k\to\infty}\sup\mathbb{E}\{\|e^x_k(t+1)\|\} \le \phi^x_t[L_{\Phi}\sigma_w + \sigma_v].\end{cases}$$

This implies that $\Phi = -[\mathbf{CB}]^{\top}\big(\mathbf{CB}[\mathbf{CB}]^{\top}\big)^{-1}\mathbf{Cf}$ can completely eliminate the influence of T on the steady-state tracking performance. When $\mathbf{Cf}$ or $\mathbf{CB}$ contains unknown information, it is difficult to estimate the value of $L^{\Phi}_{\mathbf{Cf}}$. However, due to the continuous

dependence of system tracking performance on Φ, we have the following design criteria of Φ for good steady-state tracking performance. Suppose that $\mathbf{Cf}$ can be decomposed into $\mathbf{Cf} = \widetilde{\mathbf{Cf}} + \overline{\mathbf{Cf}}$, where $\widetilde{\mathbf{Cf}}$ is unknown and $\overline{\mathbf{Cf}}$ is known. When $\mathbf{CB}$ is known and $\|\widetilde{\mathbf{Cf}}\| \ll \|\overline{\mathbf{Cf}}\|$, $\Phi = -[\mathbf{CB}]^{\top}\left(\mathbf{CB}[\mathbf{CB}]^{\top}\right)^{-1}\overline{\mathbf{Cf}}$ may be a good choice for good steady-state tracking performance. Furthermore, suppose that $\mathbf{CB}$ can be decomposed into $\mathbf{CB} = \widetilde{\mathbf{CB}} + \overline{\mathbf{CB}}$, where $\widetilde{\mathbf{CB}}$ is unknown and $\overline{\mathbf{CB}}$ is known. When $\|\widetilde{\mathbf{Cf}}\| \ll \|\overline{\mathbf{Cf}}\|$ and $\|\widetilde{\mathbf{CB}}\| \ll \|\overline{\mathbf{CB}}\|$, $\Phi = -[\overline{\mathbf{CB}}]^{\top}\left(\overline{\mathbf{CB}}[\overline{\mathbf{CB}}]^{\top}\right)^{-1}\overline{\mathbf{Cf}}$ may be a good choice for good steady-state tracking performance.

On the basis of Theorem 1, we have the following theorem, which shows how the state feedback Φ shapes transient tracking performance of output of learning system (1)–(5) with local Lipschitz nonlinearity.

Theorem 3. *Consider learning system (1)–(5), where*
(a3) $\mathbf{f}$ and Φ are locally Lipschitz continuous;
(b3) $\rho(I_q - \mathbf{CB}\Gamma) < 1$ and $q = p$;
(c3) Assumptions 2 and 3 hold.
Then, for any $y_d(\cdot)$ and $u_1(\cdot)$,

$$\mathbb{E}\{\|e_k^y(t)\|\} \le \|\Xi^{-1}\|\Psi_t \ (\forall k \in \mathbb{N}), \tag{45}$$

where $t \in \mathbb{S}^+$,

$$\Psi_1 = \max\left\{\|\Xi e_1^y(1)\|, \frac{\|\Xi\|\|\mathbf{B}\|}{1-\sigma}\|\mathbf{C}\|\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma_v\right]\right\},$$

$$\Psi_t = \max\left\{\|\Xi e_1^y(t)\|, \theta_t\right\} \text{ for } 2 \le t \le T,$$

$$\theta_t = \frac{\|\Xi\|\|\mathbf{B}\|}{1-\sigma}\|\Gamma\|\|\mathbf{H}^{-1}\|L_{\mathbf{Cf}}^{\Phi}\left[\sum_{\mu=1}^{t-1}[L_{\mathbf{f}}^{\Phi}]^{t-1-\mu}\Psi_\mu\right] + \frac{\|\Xi\|\|\mathbf{B}\|}{1-\sigma}\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma_v\right]M_{t-1} \text{ for } 2 \le t \le T,$$

$$M_t = \left[\sum_{\mu=1}^{t}(L_{\mathbf{f}}^{\Phi})^{t-\mu}\right]L_{\mathbf{Cf}}^{\Phi} + \|\mathbf{C}\| \text{ for } 1 \le t \le T-1,$$

where Ξ is the nonsingular matrix satisfying

$$\sigma = \|\Xi[I_q - \mathbf{CB}\Gamma]\Xi^{-1}\| < 1$$

and the locally Lipschitz constants $L_{\mathbf{Cf}}^{\Phi}$, $L_{\mathbf{f}}^{\Phi}$ and L_Φ are given in (25), (27) and (28), respectively.

Proof. Let $y_d(\cdot)$ and $u_1(\cdot)$ be arbitrarily given.

According to (b3) and property **P1**, there is a unique pair $\langle u_d(\cdot), x_d(\cdot)\rangle$ satisfying (2). And the boundedness of $y_d(\cdot)$, $u_d(\cdot)$, $x_d(\cdot)$ and $u_1(\cdot)$ is obvious.

Clearly, under the conditions $(a3)$, $(b3)$ and $(c3)$, Theorems 1 and 2 hold. According to Theorem 1, there is a positive constant K independent of the distributions of measurement noises such that $\forall t$ and k,

$$\|x_d(t)\| \le K, \ \|x_k(t)\| \le K \text{ and } \|\hat{x}_k(t)\| \le K.$$

Since $\mathbf{f}$ and Φ are locally Lipschitz continuous, there are nonnegative constants $L_{\mathbf{Cf}}^{\Phi}$, $L_{\mathbf{f}}$, $L_{\mathbf{f}}^{\Phi}$ and L_Φ independent of the distributions of measurement noises such that (25), (27) and (28) hold for all $z_1, z_2 \in \{z : \|z\| \le K\}$ and $t \in \mathbb{S}^-$.

Then, by (11), (25), (28) and (29) together with a simple manipulation, one has that $\forall t \in \mathbb{S}^-$,

$$\mathbb{E}\{\|\Xi e_{k+1}^y(t+1)\|\} \le \sigma\mathbb{E}\{\|\Xi e_k^y(t+1)\|\} + \|\Xi\|L_{\mathbf{Cf}}^{\Phi}\mathbb{E}\{\|\Delta x_k(t)\|\} + \|\Xi\|\|\mathbf{C}\|\|\mathbf{B}\|\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma_v\right], \tag{46}$$

where $\sigma_w = \sum_{i=1}^{n}\sigma_{i,w}$, $\sigma_v = \sum_{i=1}^{q}\sigma_{i,v}$ and Ξ is the nonsingular matrix satisfying $\sigma = \|\Xi\left[I_q - \mathbf{CB}\Gamma\right]\Xi^{-1}\| < 1$. Using (1), (5), (27) and (28) together with a simple manipulation gives

$$\mathbb{E}\{\|\Delta x_k(t)\|\} \le \|\mathbf{B}\|\|\Gamma\|\|\Xi^{-1}\|\mathbb{E}\{\|\Xi e_k^y(t)\|\} + L_{\mathbf{f}}^{\Phi}\mathbb{E}\{\|\Delta x_k(t-1)\|\} + \|\mathbf{B}\|\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma_v\right] \ (\forall t \in \mathbb{S}^+). \tag{47}$$

In (46), letting $t = 0$ and taking $x_k(0) = x^0(\forall k)$ into consideration give

$$\mathbb{E}\{\|\Xi e_{k+1}^y(1)\|\} \le \sigma\mathbb{E}\{\|\Xi e_k^y(1)\|\} + \|\Xi\|\|\mathbf{C}\|\|\mathbf{B}\|\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma_v\right]. \tag{48}$$

Moreover, for $1 \le t \le T-1$, backward iteration of (47) along time axis t and inserting it into (46) yield that for all k,

$$\mathbb{E}\{\|\Xi e_{k+1}^y(t+1)\|\} \le \sigma\mathbb{E}\{\|\Xi e_k^y(t+1)\|\} + \|\Xi\|\|\mathbf{B}\|\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma_v\right]M_t + \|\Xi\|\|\mathbf{B}\|\|\Gamma\|\|\Xi^{-1}\|L_{\mathbf{Cf}}^{\Phi}\left[\sum_{\mu=1}^{t}(L_{\mathbf{f}}^{\Phi})^{t-\mu}\mathbb{E}\{\|e_k^y(\mu)\|\}\right]. \tag{49}$$

Now, we use mathematical induction to prove

$$\mathbb{E}\{\|\Xi e_k^y(t)\|\} \le \Psi_t \ (\forall t \in \mathbb{S}^+, k \in \mathbb{N}). \tag{50}$$

Using (48) and Lemma 1 yields

$$\mathbb{E}\{\|\Xi e_k^y(1)\|\} \le \max\left\{\|\Xi e_1^y(1)\|, \frac{\|\Xi\|\|\mathbf{C}\|\|\mathbf{B}\|}{1-\sigma}\left[2L_\Phi\sigma_w + \|\Gamma\|\sigma\right]\right\} \ (\forall k \in \mathbb{N}).$$

Therefore, (50) is true for $t = 1$.

Now, we prove $\mathbb{E}\{\|\Xi e_k^y(s+1)\|\} \le \Psi_{s+1}$ provided that $\mathbb{E}\{\|H\delta y_k(t)\|\} \le \Psi_t(\forall k)$, where $1 \le t \le s < T$.

In (49), letting $t = s+1$ and taking the induction hypothesis and Lemma 1 into consideration yield

$$\mathbb{E}\{\|\Xi e_{k+1}^y(s+1)\|\} \le \left\{\|\Xi e_1^y(s+1)\|, \theta_{s+1}\right\} \ (\forall k),$$

which implies (50) holds for $t = s+1$. It follows from mathematical induction that (50) holds.

Notice that

$$\mathbb{E}\{\|e_k^y(t)\|\} \le \|\Xi^{-1}\|\mathbb{E}\{\|\Xi e_k^y(t)\|\} \ (\forall t, k).$$

Then, by (50) we have (45). □

Remark 10. Theorem 3 shows the influence factors on transient tracking performance. When **Cf** and **CB** are known, letting $\Phi = -[\mathbf{CB}]^\top\left(\mathbf{CB}[\mathbf{CB}]^\top\right)^{-1}\mathbf{Cf}$ yields $L_{\mathbf{Cf}}^{\Phi} = 0$. Let the learning gain matrix satisfy $\|I_q - CB\Gamma\| < 1$. Then, (45) becomes

$$\mathbb{E}\{\|e_k^y(t)\|\} \le \max\left\{\|e_1^y(t)\|, \frac{2\|\mathbf{C}\|\|\mathbf{B}\|L_{\Phi}}{1-\sigma}\sigma_w + \frac{\|\mathbf{C}\|\|\mathbf{B}\|\|\Gamma\|}{1-\sigma}\sigma_v\right\}.$$

This implies that $\Phi = -[\mathbf{CB}]^\top\left(\mathbf{CB}[\mathbf{CB}]^\top\right)^{-1}\mathbf{Cf}$ can completely eliminate the influence of T on the transient tracking performance and prevent ILC system (1)–(5) from experiencing overshoot. In (49), letting $\Phi = 0$ yields

$$\mathbb{E}\{\|e_{k+1}^y(t)\|\} \le \|I_q - \mathbf{CB}\Gamma\|\mathbb{E}\{\|e_k^y(t)\|\} + \|\mathbf{C}\|\|\mathbf{B}\|\|\Gamma\|\left[\sum_{\mu=1}^{t-1}(L_{\mathbf{f}})^{t-\mu}\mathbb{E}\{\|e_k^y(\mu)\|\}\right] + \|\mathbf{C}\|\|\mathbf{B}\|\|\Gamma\|\left[\sum_{s=1}^{t}(L_{\mathbf{f}})^s\right]\sigma_v (2 \le t \le T),$$

which reveals how **f** shapes the expected transient tracking performance of output and how T affects the expected transient tracking performance when $L_{\mathbf{f}} > 1$. When **f**, **Cf** or **CB** contains unknown information, it is difficult to estimate the values of $L_{\mathbf{f}}$ and $L_{\mathbf{Cf}}^{\Phi}$. According to the continuous dependence of transient tracking performance of ILC system (1)–(5) on Φ, we have the following design criteria of Φ for good transient tracking performance. When **CB** is known and $\|\widetilde{\mathbf{Cf}}\| \ll \|\overline{\mathbf{Cf}}\|$, $\Phi = -[\mathbf{CB}]^\top\left(\mathbf{CB}[\mathbf{CB}]^\top\right)^{-1}\overline{\mathbf{Cf}}$ may be a good choice for good transient tracking performance, where $\mathbf{Cf} = \widetilde{\mathbf{Cf}} + \overline{\mathbf{Cf}}$, $\widetilde{\mathbf{Cf}}$ is unknown and $\overline{\mathbf{Cf}}$ is known. When $\|\widetilde{\mathbf{Cf}}\| \ll \|\overline{\mathbf{Cf}}\|$ and $\|\widetilde{\mathbf{CB}}\| \ll \|\overline{\mathbf{CB}}\|$, $\Phi = -[\overline{\mathbf{CB}}]^\top\left(\overline{\mathbf{CB}}[\overline{\mathbf{CB}}]^\top\right)^{-1}\overline{\mathbf{Cf}}$ may be a good choice for good transient tracking performance, where $\mathbf{CB} = \widetilde{\mathbf{CB}} + \overline{\mathbf{CB}}$, $\widetilde{\mathbf{CB}}$ is unknown and $\overline{\mathbf{CB}}$ is known.

4. Convergence results for the case $q < p$

In this section, we use the input transform technique to investigate the robust convergence under $\text{rank}(CB) = q < p$. The following theorem presents the bounded convergence results under the condition **f** and Φ are continuous.

Theorem 4. *Consider learning system (1)–(5), where*

(a4) **f** *and* Φ *are continuous;*

(b4) $\rho(I_q - \mathbf{CB}\Gamma) < 1$ *and* $q < p$*;*

(c4) Assumption 2 holds.

Then, for any $y_d(\cdot)$, $u_1(\cdot)$ *and* $t(\in \mathbb{S}^-)$*, there are positive real numbers* C_t^y, C_t^u *and* C_t^x *independent of the distributions of measurement noises such that*

$$\|e_k^y(t+1)\| \le C_t^y, \|e_k^u(t)\| \le C_t^u \text{ and } \|e_k^x(t+1)\| \le C_t^x \tag{51}$$

$\forall k \in \mathbb{N}$*, where* $x_d(\cdot)$ *is the limit of* $x_k(\cdot)$ *and* $u_d(\cdot)$ *is the limit of* $u_k(\cdot)$ *in the absence of measurement noises;* $x_d(\cdot)$, $u_d(\cdot)$ *and* $y_d(\cdot)$ *satisfy Eq. (2).*

Remark 11. The condition (b4) implies $\text{rank}(\mathbf{CB}) = q < p$. However, we cannot directly use the proof method of Theorem 1 to prove Theorem 4. We will show that by introducing input transform technique, ILC system (1)–(5) can be equivalently transformed into a special ILC system with $p - q$ components being iteration-invariant.

Proof. Let $y_d(\cdot)$ and $u_1(\cdot)$ be arbitrarily given, respectively. Clearly, $y_d(\cdot)$ and $u_1(\cdot)$ are bounded.

According to Lemma 2, for learning system (1)–(5) without measurement noises, the condition $\rho(I_q - \mathbf{CB}\Gamma) < 1$ implies that $\lim_{k\to\infty} y_k(\cdot) = y_d(\cdot)$ and $x_k(\cdot)$ and $u_k(\cdot)$ converge along the k axis. Let

$$u_d(\cdot) = \lim_{k\to\infty} u_k(\cdot) \text{ and } x_d(\cdot) = \lim_{k\to\infty} x_k(\cdot).$$

$y_d(\cdot)$, $u_d(\cdot)$ and $x_d(\cdot)$ satisfy Eq. (2). Clearly, $u_d(\cdot)$ and $x_d(\cdot)$ are bounded.

Using (b4), we have $\text{rank}(\Gamma) = q < p$. Therefore, we can find an invertible matrix Ξ such that

$$\Xi\Gamma = \begin{bmatrix} I_q \\ 0_{(p-q)\times q} \end{bmatrix}. \tag{52}$$

Define

$$\bar{u}_d(t) = \Xi\left[u_d(t) - \boldsymbol{\Phi}(t, x_d(t))\right].$$

Partition $\mathbf{B}\Xi^{-1}$ and $\bar{u}_d(t)$ into

$$\mathbf{B}\Xi^{-1} = \begin{bmatrix}\bar{\mathbf{B}}_1 & \bar{\mathbf{B}}_2\end{bmatrix} \text{ and } \bar{u}_d(t) = \begin{bmatrix}\bar{u}_d^1(t)\\ \bar{u}_d^2(t)\end{bmatrix}, \tag{53}$$

where $\bar{u}_d^1(t) \in \mathbb{R}^q$, $\bar{u}_d^2(t) \in \mathbb{R}^{p-q}$, $\bar{\mathbf{B}}_1 \in \mathbb{R}^{n\times q}$ and $\bar{\mathbf{B}}_2 \in \mathbb{R}^{n\times(p-q)}$. Then, Eq. (2) can be split into

$$\begin{cases} x_d(t+1) = \mathbf{f}(t, x_d(t)) + \mathbf{B}\boldsymbol{\Phi}(t, x_d(t)) + \bar{\mathbf{B}}_1\bar{u}_d^1(t) + \bar{\mathbf{B}}_2\bar{u}_d^2(t), t \in \mathbb{S}^-,\\ y_d(t) = \mathbf{C}x_d(t), t \in \mathbb{S}^+,\\ x_d(0) = x^0. \end{cases} \tag{54}$$

Define

$$\bar{u}_k(t) = \Xi[u_k(t) - \boldsymbol{\Phi}(t, \hat{x}_k(t))].$$

Partition $\bar{u}_k(t)$ into

$$\bar{u}_k(t) = \begin{bmatrix}\bar{u}_k^1(t)\\ \bar{u}_k^2(t)\end{bmatrix}, \tag{55}$$

where $\bar{u}_k^1(t) \in \mathbb{R}^q$ and $\bar{u}_k^2(t) \in \mathbb{R}^{p-q}$. Then, (5) becomes

$$\bar{u}_k^2(t) = \bar{u}_1^2(t) \text{ for } t \in \mathbb{S}^-, \tag{56}$$

$$\bar{u}_{k+1}^1(t) = \bar{u}_k^1(t) + e_k^y(t+1) - \upsilon_k(t+1) \text{ for } t \in \mathbb{S}^- \tag{57}$$

and system (1) becomes

$$\begin{cases} x_k(t+1) = \mathbf{f}(t, x_k(t)) + \mathbf{B}\boldsymbol{\Phi}(t, \hat{x}_k(t)) + \bar{\mathbf{B}}_1\bar{u}_k^1(t) + \bar{\mathbf{B}}_2\bar{u}_1^2(t),\\ y_k(t) = \mathbf{C}x_k(t),\\ x_k(0) = x^0. \end{cases} \tag{58}$$

It is observed from (56) that under the input transform, $\bar{u}_k^2(\cdot)$ is iteration-invariant regardless of the presence or absence of measurement noises. Accordingly,

$$\bar{u}_d^2(t) = \bar{u}_1^2(t) \ (t \in \mathbb{S}^-). \tag{59}$$

Using (52) and (53) together with a direct computation yields

$$CB\Gamma = C\bar{B}_1 \text{ and } B\Gamma = \bar{B}_1. \tag{60}$$

Due to $\rho(I_q - CB\Gamma) < 1$, $CB\Gamma$ is nonsingular. It follows from (60) that $C\bar{B}_1$ is nonsingular.

Next, we first show that for any $t \in \mathbb{S}^-$, there are positive real number $\bar{C}_t^y$, $\bar{C}_t^u$ and $\bar{C}_t^x$ independent of the distributions of measurement noises such that $\forall t \in \mathbb{S}^-$,

$$\|e_k^y(t+1)\| \le \bar{C}_t^y, \|\delta\bar{u}_k^1(t)\| \le \bar{C}_t^y \text{ and } \|e_k^x(t+1)\| \le \bar{C}_t^x, \tag{61}$$

where $\delta\bar{u}_k^1(t) = \bar{u}_d^1(t) - \bar{u}_k^1(t)$.

Using (54), (57), (58) and (60) yields

$$e_{k+1}^y(t+1) = [I_q - \mathbf{CB}\Gamma]e_k^y(t+1) + \delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{CB}}(t, x_{k+1}(t)) - \mathbf{CB}\left[\Delta_w\boldsymbol{\Phi}(t, x_{k+1}(t)) - \Delta_w\boldsymbol{\Phi}(t, x_k(t))\right] - \delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{CB}}(t, x_k(t)) + \mathbf{CB}\Gamma\upsilon_k(t+1) \tag{62}$$

and

$$\begin{cases} e_k^x(t+1) = \delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{B}}(t, x_k(t)) - \mathbf{B}\Delta_w\boldsymbol{\Phi}(t, x_k(t)) + \bar{\mathbf{B}}_1\delta\bar{u}_k^1(t),\\ \delta\bar{u}_k^1(t) = [\mathbf{CB}\Gamma]^{-1}\left[e_k^y(t+1) - \delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{CB}}(t, x_k(t)) + \mathbf{CB}\Delta_w\boldsymbol{\Phi}(t, x_k(t))\right]. \end{cases} \tag{63}$$

Since $\rho(I_q - \mathbf{CB}\Gamma) < 1$, we can find a nonsingular matrix $\Xi \in \mathbb{R}^{q\times q}$ such that $\sigma = \|\Xi[I_q - \mathbf{CB}\Gamma]\Xi^{-1}\| < 1$. By (62) and (63) we obtain

$$\begin{aligned}\|\Xi e_{k+1}^y(t+1)\| \le{}& \|\Xi\mathbf{CB}\|\left[\|\Delta_w\boldsymbol{\Phi}(t, x_{k+1}(t))\| + \|\Delta_w\boldsymbol{\Phi}(t, x_k(t))\|\right] + \|\Xi\|\left[\|\delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{CB}}(t, x_{k+1}(t))\| + \|\delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{CB}}(t, x_k(t))\|\right]\\ &+ \sigma\|\Xi e_k^y(t+1)\| + \|\Xi\mathbf{CB}\Gamma\|\|\upsilon_k(t+1)\|\end{aligned} \tag{64}$$

and

$$\begin{cases} \|e_k^x(t+1)\| \le \|\delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{B}}(t, x_k(t))\| + \|\bar{\mathbf{B}}_1\|\|\delta\bar{u}_k^1(t)\| + \|\mathbf{B}\|\|\Delta_w\boldsymbol{\Phi}(t, x_k(t))\|,\\ \|\delta\bar{u}_k^1(t)\| \le \|[\mathbf{CB}\Gamma]^{-1}\|\left[\|e_k^y(t+1)\| + \|\delta\mathbf{f}_{\boldsymbol{\Phi}}^{\mathbf{CB}}(t, x_k(t))\| + \|\mathbf{CB}\Delta_w\boldsymbol{\Phi}(t, x_k(t))\|\right]. \end{cases} \tag{65}$$

Notice $\|v_k(t+1)\| < qC_v$ and $\|w_k(t)\| \le nC_w$ $\forall t,k$. Then, using (64) and (65) and following the same technical route of the proof of Theorem 1 gives (61). For saving space, we omit it.

Notice that $\|\delta\bar{u}_k(t)\| = \|\delta\bar{u}_k^1(t)\|$ and

$$e_k^u(t) = \Xi^{-1}\left[\bar{u}_d(t) - \bar{u}_k(t)\right] + \left[\boldsymbol{\Phi}(t, x_d(t)) - \boldsymbol{\Phi}(t, \hat{x}_k(t))\right]$$

$\forall t \in \mathbb{S}^-$. Taking the continuity of $\boldsymbol{\Phi}$ and (61) into consideration yields (51). □

The following theorem gives the relations between the steady-state tracking performance and the standard deviations σ_v and σ_w under the condition f and $\boldsymbol{\Phi}$ are locally Lipschitz continuous.

Theorem 5. *Consider learning system (1)–(5), where*

(a5) $\mathbf{f}$ *and* $\boldsymbol{\Phi}$ *are locally Lipschitz continuous;*

(b5) $\rho(I_q - \mathbf{CB}\Gamma) < 1$ *and* $q < p$;

(c5) Assumptions 2 and 3 hold.

Then, for any $y_d(\cdot)$ *and* $u_1(\cdot)$, *there exist the nonnegative constants* ϕ_t^y, ϕ_t^u *and* ϕ_t^x $(t \in \mathbb{S}^-)$, *independent of* $\boldsymbol{\Phi}$ *and the measurement noises such that* $\forall t \in \mathbb{S}^-$,

$$\begin{cases} \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^y(t+1)\|\} \le \phi_t^y \left\{ \left[\sum\limits_{s=0}^{t} (L_{\mathbf{Cf}}^{\boldsymbol{\Phi}})^s\right] + \left[\sum\limits_{s=0}^{t} (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^s\right] \right\} F(\sigma_w, \sigma_v), \\ \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^u(t)\|\} \le \phi_t^u \left\{ \left[\sum\limits_{s=0}^{t} (L_{\mathbf{Cf}}^{\boldsymbol{\Phi}})^s\right] + \left[\sum\limits_{s=0}^{t} (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^s\right] \right\} F(\sigma_w, \sigma_v), \\ \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^x(t+1)\|\} \le \phi_t^x \left\{ \left[\sum\limits_{s=0}^{t} (L_{\mathbf{Cf}}^{\boldsymbol{\Phi}})^s\right] + \left[\sum\limits_{s=0}^{t} (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^s\right] \right\} F(\sigma_w, \sigma_v), \end{cases} \tag{66}$$

where $\sigma_w = \sum_{i=1}^{n} \sigma_{i,w}$ *and* $\sigma_v = \sum_{i=1}^{q} \sigma_{i,v}$; $x_d(\cdot)$ *and* $u_d(\cdot)$ *are the limits of* $x_k(\cdot)$ *and* $u_k(\cdot)$ *in the absence of measurement noises;* $y_d(\cdot)$, $u_d(\cdot)$ *and* $x_d(\cdot)$ *satisfy Eq. (2).*

Remark 12. Since the estimations of the upper bound of the steady-state tracking performance of learning system (1)–(5) for the case $\mathrm{rank}(\mathbf{CB}) = q < p$ include the factor $1 + L_{\mathbf{f}}^{\boldsymbol{\Phi}} + \cdots + (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^t$, unlike the case $\mathrm{rank}(CB) = q = p$, we cannot judge whether state feedback can theoretically reduce the effect of large trial length on steady-state tracking performance.

Proof. Let $y_d(\cdot)$ and $u_1(\cdot)$ be arbitrarily given, respectively.

Under the conditions (a5), (b5) and (c5), Theorem 4 holds, which implies that under the same manipulation, all the relations given in Theorem 4 hold. It follows from Theorem 4 that $y_d(\cdot)$, $u_d(\cdot)$, $x_d(\cdot)$, $y_k(\cdot)$, $u_k(\cdot)$ and $x_k(\cdot)$ are bounded, and there exists a positive constant K independent of the distributions of measurement noises such that $\max_{t,k} \|x_d(t)\| \le K$, $\max_{t,k} \|x_k(t)\| \le K$ and $\max_{t,k} \|\hat{x}_k(t)\| \le K$. According to (a5), there are nonnegative constants $L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}$, $L_{\mathbf{f}}$, $L_{\mathbf{f}}^{\boldsymbol{\Phi}}$ and $L_{\boldsymbol{\Phi}}$, independent of the distributions of measurement noises, satisfy the inequalities (25), (26), (27), and (28), respectively. Then, using (64), (29) and (65) together with a simple manipulation gives

$$\mathbb{E}\{\|\Xi e_{k+1}^y(t+1)\|\} \le \sigma \mathbb{E}\{\|\Xi e_k^y(t+1)\|\} + 2\|\Xi\mathbf{CB}\|L_{\boldsymbol{\Phi}}\sigma_w + \|\Xi\mathbf{CB}\Gamma\|\sigma_v + \|\Xi\|L_{\mathbf{Cf}}^{\boldsymbol{\Phi}}\left[\mathbb{E}\{\|e_{k+1}^x(t)\|\} + \mathbb{E}\{\|e_k^x(t)\|\}\right] \tag{67}$$

and

$$\begin{cases} \mathbb{E}\{\|e_k^x(t+1)\|\} \le L_{\mathbf{f}}^{\boldsymbol{\Phi}}\mathbb{E}\{\|e_k^x(t)\|\} + \|\bar{\mathbf{B}}_1\|\mathbb{E}\{\|\delta\bar{u}_k^1(t)\|\} + \|\mathbf{B}\|L_{\boldsymbol{\Phi}}\sigma_w, \\ \mathbb{E}\{\|\delta\bar{u}_k^1(t)\|\} \le \|[CB\Gamma]^{-1}\| \left[\, \mathbb{E}\{\|e_k^y(t+1)\|\} + L_{\mathbf{f}}^{\boldsymbol{\Phi}}\mathbb{E}\{\|e_k^x(t)\|\} + \|\mathbf{CB}\|L_{\boldsymbol{\Phi}}\sigma_w \,\right], \end{cases} \tag{68}$$

where Ξ is a nonsingular matrix such that $\sigma = \|\Xi[I_q - \mathbf{CB}\Gamma]\Xi^{-1}\| < 1$.

Using (67) and (68) together with the same technique route as that used in the proof of Theorem 2, it is easy to prove that there are nonnegative constants $\bar{\phi}_t^y$, $\bar{\phi}_t^u$ and $\bar{\phi}_t^x$ $(t \in \mathbb{S}^-)$, independent of $\boldsymbol{\Phi}$ and the measurement noises, such that $\forall t \in \mathbb{S}^-$,

$$\begin{cases} \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^y(t+1)\|\} \le \bar{\phi}_t^y \left\{ \left[\sum\limits_{s=0}^{t} (L_{\mathbf{Cf}}^{\boldsymbol{\Phi}})^s\right] + \left[\sum\limits_{s=0}^{t} (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^s\right] \right\} F(\sigma_w, \sigma_v), \\ \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|\delta\bar{u}_k^1(t)\|\} \le \bar{\phi}_t^u \left\{ \left[\sum\limits_{s=0}^{t} (L_{\mathbf{Cf}}^{\boldsymbol{\Phi}})^s\right] + \left[\sum\limits_{s=0}^{t} (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^s\right] \right\} F(\sigma_w, \sigma_v), \\ \lim\limits_{k\to\infty} \sup \mathbb{E}\{\|e_k^x(t+1)\|\} \le \bar{\phi}_t^x \left\{ \left[\sum\limits_{s=0}^{t} (L_{\mathbf{Cf}}^{\boldsymbol{\Phi}})^s\right] + \left[\sum\limits_{s=0}^{t} (L_{\mathbf{f}}^{\boldsymbol{\Phi}})^s\right] \right\} F(\sigma_w, \sigma_v). \end{cases} \tag{69}$$

It is easy to check that

$$e_k^u(t) = \Xi^{-1}\delta\bar{u}_k(t) + [\boldsymbol{\Phi}(t, x_d(t)) - \boldsymbol{\Phi}(t, \hat{x}_k(t))] \ (\forall t \in \mathbb{S}^-). \tag{70}$$

Notice that $\|\delta\bar{u}_k(t)\| = \|\delta\bar{u}_k^1(t)\| (\forall t \in \mathbb{S}^-)$. Then, by (28), (29) and (70) together with a simple manipulation, we have

$$\mathbb{E}\{\|e_k^u(t)\|\} \le \|\Xi^{-1}\|\mathbb{E}\{\|\delta\bar{u}_k(t)\|\} + L_{\boldsymbol{\Phi}}\mathbb{E}\{\|e_k^x(t)\|\} + L_{\boldsymbol{\Phi}}\sigma_w, \forall t \in \mathbb{S}^-. \tag{71}$$

By (69) and (71), we can find nonnegative constants ϕ_t^y, ϕ_t^u and ϕ_t^x ($t \in \mathbb{S}^-$), independent of Φ and the measurement noises such that (66) holds for all $t \in \mathbb{S}^-$. □

The following theorem establishes the relation between the output transient tracking performance and the local Lipschitz constants $L_{\mathbf{Cf}}^{\Phi}$, $L_{\mathbf{f}}^{\Phi}$ and Φ, which reveals how Φ shapes the transient tracking performance of ILC system (1)–(5) under the condition f and Φ are locally Lipschitz continuous.

Theorem 6. *Consider learning system (1)–(5), where*
(a6) $\mathbf{f}$ *and* Φ *are locally Lipschitz continuous;*
(b6) $\rho(I_q - \mathbf{CB}\Gamma) < 1$ *and* $q < p$*;*
(c6) Assumptions 2 and 3 hold.
Then, for any $y_d(\cdot)$ *and* $u_1(\cdot)$*,*

$$\mathbb{E}\{\|e_k^y(t)\|\} \le \|\Xi^{-1}\|\Psi_t \ (\forall t \in \mathbb{S}^+,\ k \in \mathbb{N}), \tag{72}$$

where

$$\Psi_1 = \max\left\{\|\Xi e_1^y(1)\|, \frac{\|\Xi\|\|\mathbf{B}\|}{1-\sigma}\|\mathbf{C}\|\left[2L_{\Phi}\sigma_w + \|\Gamma\|\sigma_v\right]\right\},$$

$$\Psi_t = \max\left\{\|\Xi e_1^y(t)\|, \theta_t\right\} \text{ for } 2 \le t \le T,$$

$$\theta_t = \frac{\|\Xi\|\|\mathbf{B}\|}{1-\sigma}\|\Gamma\|\|\Xi^{-1}\|L_{\mathbf{Cf}}^{\Phi}\left[\sum_{\mu=1}^{t-1}[L_{\mathbf{f}}^{\Phi}]^{t-1-\mu}\Psi_{\mu}\right] + \frac{\|\Xi\|\|\mathbf{B}\|}{1-\sigma}\left[2L_{\Phi}\sigma_w + \|\Gamma\|\sigma_v\right]M_{t-1} \text{ for } 2 \le t \le T,$$

$$M_t = \left[\sum_{\mu=1}^{t}(L_{\mathbf{f}}^{\Phi})^{t-\mu}\right]L_{\mathbf{Cf}}^{\Phi} + \|\mathbf{C}\| \text{ for } 1 \le t \le T-1,$$

where Ξ *is a nonsingular matrix satisfying*

$$\|\Xi[I_q - \mathbf{CB}\Gamma]\Xi^{-1}\| < 1$$

and $L_{\mathbf{Cf}}^{\Phi}$, $L_{\mathbf{f}}^{\Phi}$ *and* L_{Φ} *are given in (25), (27) and (28), respectively.*

Proof. Let $y_d(\cdot)$ and $u_1(\cdot)$ be arbitrarily given, respectively.

Under the conditions (a6), (b6) and (c6), Theorems 4 and 5 hold, which imply that under the same manipulation, all the relations given in Theorems 4 and 5 hold. It follows from Theorem 4 that $y_d(\cdot)$, $u_d(\cdot)$, $x_d(\cdot)$, $y_k(\cdot)$, $u_k(\cdot)$ and $x_k(\cdot)$ are bounded, and there exists a positive constant K, independent of the distributions of measurement noises, such that $\max_{t,k}\|x_d(t)\| \le K$, $\max_{t,k}\|x_k(t)\| \le K$ and $\max_{t,k}\hat{x}_k(t)$. According to (a6), there are nonnegative constants $L_{\mathbf{Cf}}^{\Phi}$, $L_{\mathbf{f}}$, $L_{\mathbf{f}}^{\Phi}$ and L_{Φ}, independent of the distributions of measurement noises, that satisfy the inequalities (25), (26), (27), and (28), respectively. Then, using (57), (58), (60), (25), (27), (28) and (29) together with a simple manipulation gives

$$\mathbb{E}\{\|\Xi e_{k+1}^y(t+1)\|\} \le \sigma\mathbb{E}\{\|\Xi e_k^y(t+1)\|\} + \|\Xi\|L_{\mathbf{Cf}}^{\Phi}\mathbb{E}\{\|\Delta x_k(t)\|\} + \|\Xi\|\|\mathbf{C}\|\|\mathbf{B}\|\left[2L_{\Phi}\sigma_w + \|\Gamma\|\sigma_v\right] \tag{73}$$

and

$$\mathbb{E}\{\|\Delta x_k(t+1)\|\} \le L_{\mathbf{f}}^{\Phi}\mathbb{E}\{\|\Delta x_k(t)\|\} + \|\bar{\mathbf{B}}_1\|\|\Xi^{-1}\|\mathbb{E}\{\|\Xi e_k^y(t+1)\|\} + 2\|\mathbf{B}\|L_{\Phi}\sigma_w + \|\bar{B}_1\|\sigma_v \ (\forall t \in \mathbb{S}^-, k \in \mathbb{N}), \tag{74}$$

where $\sigma = \|\Xi[I_q - \mathbf{CB}\Gamma]\Xi^{-1}\|$.

Using (52) and (53) gives $\bar{\mathbf{B}}_1 = \mathbf{B}\Gamma$. Then, (74) becomes

$$\mathbb{E}\{\|\Delta x_k(t)\|\} \le L_{\mathbf{f}}^{\Phi}\mathbb{E}\{\|\Delta x_k(t-1)\|\} + \|\mathbf{B}\|\|\Gamma\|\|\Xi^{-1}\|\mathbb{E}\{\|\Xi e_k^y(t)\|\} + 2\|\mathbf{B}\|L_{\Phi}\sigma_w + \|\mathbf{B}\|\|\Gamma\|\sigma_v(\forall t \in \mathbb{S}^-). \tag{75}$$

Then, using (73) and (75), and along the same technique route as used in the proof of Theorem 3, we can prove Eq. (72). For saving space, we omit it. □

Remark 13. The main differences between the results presented in this paper and those given in [4–6] lie in two aspects: the systems targeted are different and the connotation of trajectory tracking is different. The system considered in this article operates repeatedly within a finite-time interval, while the system considered in the literature operates over an infinite-time interval. The trajectory tracking considered in this paper refers to the situation where the system achieves tracking of the desired trajectory over the whole finite time interval as the number of learning iterations approaches infinity; while the trajectory tracking considered in [4–6] refers to the situation where the system achieves tracking of the reference signal as time approaches infinity.

5. Illustrative example

This section gives an illustrative example to demonstrate the influences of nonlinear state feedback and gain matrix on transient and steady-state tracking performances and the convergence speed of the considered learning system with measurement noises.

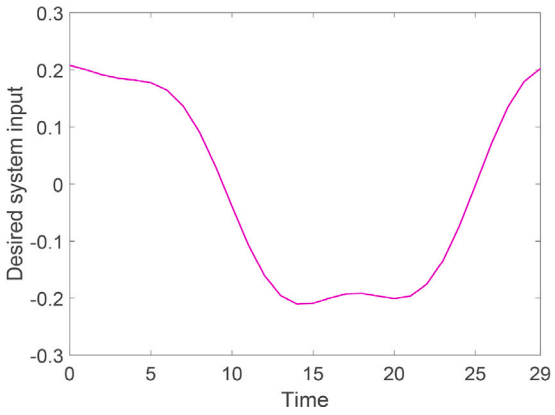

Fig. 1. Desired input $u_d(\cdot)$ corresponding to $y_d(t) = \sin(\frac{\pi t}{15})$.

Consider the discrete-time system with local Lipschitz nonlinearity taking the form of

$$\begin{cases} x_k^1(t+1) = 0.5(x_k^1(t))^2 + 0.3x_k^2(t), \\ x_k^2(t+1) = x_k^1(t) + \sin(x_k^2(t)) + u_k(t), \\ y_k(t) = x_k^2(t), \\ x_k^1(0) = x_k^2(0) = 0, \end{cases} \tag{76}$$

where $x_k^1(t)$ and $x_k^2(t)$ are the 1st and 2nd components of state $x_k(t)$, i.e., $x_k(t) = [x_k^1(t), x_k^2(t)]^\top$.

Clearly, the nonlinear function

$$\mathbf{f}(t, x_k(t)) = \begin{bmatrix} 0.5(x_k^1(t))^2 + 0.3x_k^2(t) \\ x_k^1(t) + \sin(x_k^2(t)) \end{bmatrix}$$

is locally Lipschitz continuous w.r.t. $x_k(t)$.

Due to $\mathbf{B} = [0, 1]^\top$ and $\mathbf{C} = [0, 1]$, $\mathbf{CB} = 1$, which implies that each $y_d(\cdot)$ corresponds to a unique $u_d(\cdot)$. Let $y_d(t) = \sin(\frac{\pi t}{15})$ and the trial length be $T = 30$. Clearly, the unique $u_d(\cdot)$ can be given by (3), whose profile is presented in Fig. 1.

In simulations, assume that $\boldsymbol{\Phi}$ takes the form of $\boldsymbol{\Phi} = \alpha\mathbf{Cf}$ and the output and state measurement noises $v_k(t)$, $w_k^1(t)$ and $w_k^2(t)$ are subject to the uniform distribution on the interval $[-a, a]$, respectively, i.e.,

$$v_k(t), w_k^1(t), w_k^2(t) \sim U[-a, a],$$

where the positive real a is the measurement accuracy and $\alpha \in [0, -1]$ denotes the available information of $\mathbf{Cf}$. Clearly, $|v_k(t)| \le a$, $|w_k^i(t)| \le a (i = 1, 2)$, $\mathbb{E}\{v_k(t+1)\} = \mathbb{E}\{w_k^1(t)\} = \mathbb{E}\{w_k^2(t)\} = 0$ and

$$\mathbb{E}\{[v_k(t+1)]^2\} = \mathbb{E}\{[w_k^1(t)]^2\} = \mathbb{E}\{[w_k^2(t)]^2\} = \frac{a^2}{3}$$

$\forall t \in \mathbb{S}^-,\ k \in \mathbb{N}$.

When $a = 0.01$, a direct computation yields that $\sigma_v = \sigma_{1,w} = \sigma_{2,w} = \frac{\sqrt{3}}{300}$. According to Theorems 2 and 3, when $\Gamma \in (0, 2)$, the output tracking error sequences in the sense of expectation have the bounded convergence. Since $\mathbf{f}$ and $\boldsymbol{\Phi}$ are locally Lipschitz continuous, we cannot give an accurate estimate of the upper limit of $\mathbb{E}\{|y_d(t) - y_k(t)|\}$ for given $t \in \mathbb{S}^+$.

Furthermore, we use the mean of 1000-group experiments $\frac{1}{1000}\sum_{s=1}^{1000}|y_d(t) - y_k^s(t)|$ as the approximate value of $\mathbb{E}\{|y_d(t) - y_k(t)|\}$, where $y_k^s(\cdot)$ denotes the output at the kth trial of the sth group. In simulations, we use $\max_{1\le t\le 30}\frac{1}{1000}\sum_{s=1}^{1000}|y_d(t) - y_k^s(t)|$ to evaluate the tracking performance, which is termed the maximum mean output tracking error of output at the kth trial of the 1000-group experiments.

5.1. Influence of state feedback on tracking performance

Figs. 2–4 show the influence of $\boldsymbol{\Phi}$ on MMOTE under fixed learning gain. From Figs. 2–4, one can observe that in the presence of measurement noises, $\boldsymbol{\Phi}$ can completely eliminate the overshoot and can be used to improve convergence speed and steady-state performance. However, the influence of $\boldsymbol{\Phi}$ on steady-state tracking performance and convergence speed is complicated. It is not true that the higher the information utilization, the better the steady-state tracking performance and the faster the convergence speed. Since the system considered in ILC performs repetitive tasks, the convergence speed becomes relatively less important than the transient and steady-state tracking performance. The more information of $\mathbf{Cf}$ available, the more beneficial it is to design $\boldsymbol{\Phi}$ such that learning system has a good transient tracking performance. It is still a challenging problem to design $\boldsymbol{\Phi}$ for the optimal

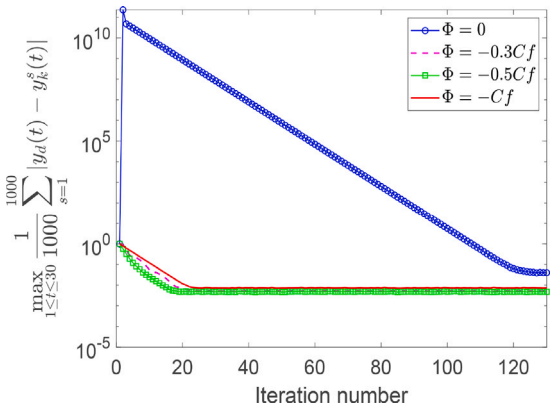

Fig. 2. Maximum mean output tracking errors under $\Gamma = 0.21$ and $\boldsymbol{\Phi} = 0, -0.3\mathbf{Cf}, -0.5\mathbf{Cf}, -\mathbf{Cf}$, respectively.

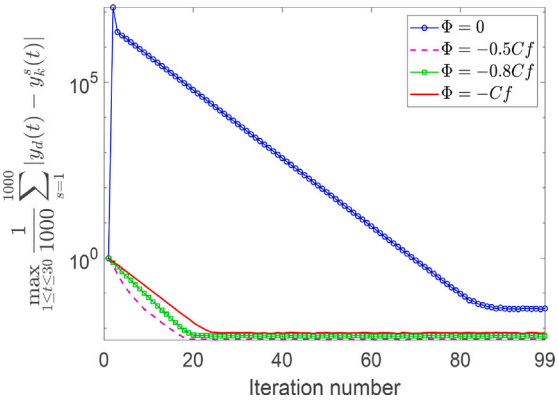

Fig. 3. Maximum mean output tracking errors under $\Gamma = 0.2$ and $\boldsymbol{\Phi} = 0, -0.5\mathbf{Cf}, -0.8\mathbf{Cf}, -\mathbf{Cf}$, respectively.

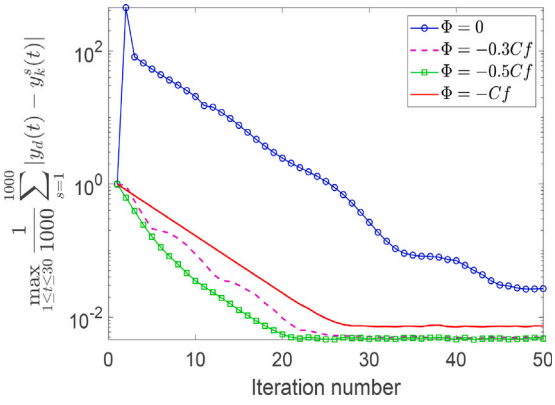

Fig. 4. Maximum mean output tracking errors under $\Gamma = 0.18$ and $\boldsymbol{\Phi} = 0, -0.3\mathbf{Cf}, -0.5\mathbf{Cf}, -\mathbf{Cf}$, respectively.

steady-state tracking performance under the premise of good transient tracking performance. In addition, as one can see from Figs. 2–4, when $\boldsymbol{\Phi} = 0$, the transient tracking performance is sensitive to the value of Γ, and state feedback mechanism can greatly reduce the sensitivity of transient and steady-state tracking performance to the learning gain.

5.2. Influence of learning gain on tracking performances

Figs. 5–10 show the influence of Γ on the tracking performance under $\boldsymbol{\Phi} = -0.8\mathbf{Cf}$ and $\boldsymbol{\Phi} = -0.5\mathbf{Cf}$, respectively. As one can see from Figs. 5 and 6, when the learning gain is greater than 1, the transient tracking performance is sensitive to Γ, and large

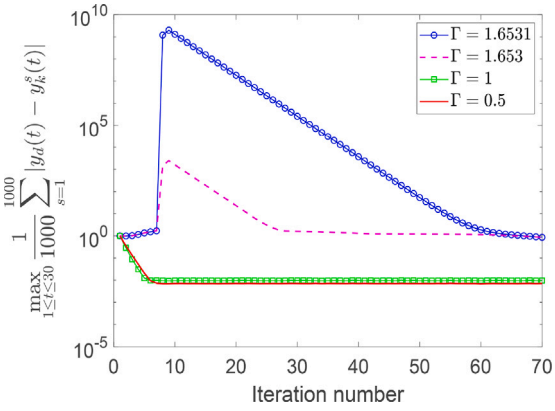

Fig. 5. Maximum mean output tracking errors under $\Phi = -0.8\mathbf{Cf}$ and $\Gamma = 1.6531, 1.653, 1, 0.5$, respectively.

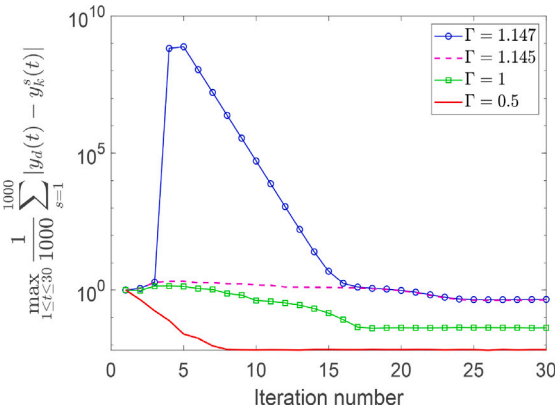

Fig. 6. Maximum mean output tracking errors under $\Phi = -0.5\mathbf{Cf}$ for $\Gamma = 1.147, 1.145, 1, 0.5$, respectively.

learning gain may mean poor transient tracking performance. It is observed from 5–8 that Γ can be used to improve steady-state tracking performance. When the learning gain is less than 1, the constant learning gain may not be able to account for steady-state tracking performance and convergence speed at the same time. Iteration-varying learning gains may be a good choice to handle this problem. Firstly, a relatively large learning gain less than or equal to 1 is used to make the learning process quickly reach a steady state, and then a relatively small learning gain is used to improve the tracking performance. In addition, as one can see from Figs. 9 and 10, it is not that the smaller the learning gain the better the steady-state tracking performance. The main reason may be that when the learning gain is very small, the measurement error accuracy and the standard deviation dominate the steady-state tracking performance.

Remark 14. According to Remarks 9 and 10, in order guarantee good steady-state and transient tracking performance of ILC system (1)–(5), the design method of Φ is to minimize the local Lipschitz constant $L_{\mathbf{Cf}}^{\Phi}$ of function $\mathbf{Cf} + \mathbf{CB}\Phi$. This implies that the design of Φ depends on $\mathbf{Cf}$ and $\mathbf{CB}$. Therefore, if $\mathbf{Cf}$ is completely unknown, we cannot design Φ taking the form of $\Phi = \alpha\mathbf{Cf}$. Since system (1) can be repeatedly operated over finite-time interval, we can obtain the information of $\mathbf{Cf}$ by using the system identification methods. Therefore, $\mathbf{Cf}$ can be decomposed into $\mathbf{Cf} = \widetilde{\mathbf{Cf}} + \overline{\mathbf{Cf}}$, where $\widetilde{\mathbf{Cf}}$ is unknown and $\overline{\mathbf{Cf}}$ is known. Consequently, $\overline{\mathbf{Cf}}$ can be used to design Φ. One purpose of simulations in this paper is to demonstrate the relationship between the amount of information of $\mathbf{Cf}$ utilized and the steady-state and transient tracking performance. However, how to describe the amount of information of $\mathbf{Cf}$ utilized when designing Φ is a scientific issue worthy of study. In simulations, for simplicity, Φ takes the form of $\Phi = \alpha\mathbf{Cf}$, where $-\alpha(\in [0, 1])$ is used to describe the amount of information of $\mathbf{Cf}$ utilized when designing Φ. The larger the value of $-\alpha$ is, the greater the amount of information of $\mathbf{Cf}$ utilized will be. So far, we still do not have a more intuitive design method of Φ for presenting the amount of information of $\mathbf{Cf}$ utilized. This is the direction we should strive for in the future.

6. Conclusions

In this paper, we investigate how state feedback shapes the tracking performance of ILC system (1)–(5) in the presence of output and state measurement noises, where the continuous nonlinearity and the local Lipschitz nonlinearity are considered successively.

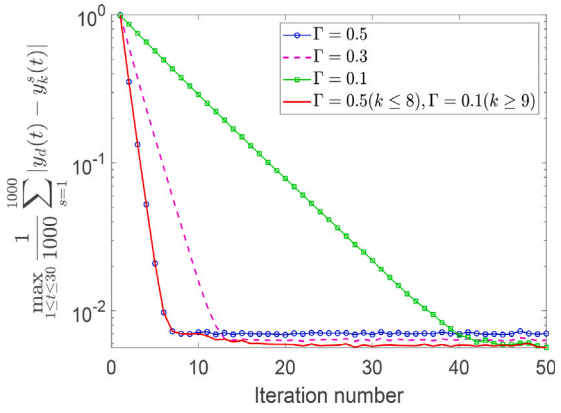

Fig. 7. Maximum mean output tracking errors for different learning gains under $\Phi = -0.8Cf$.

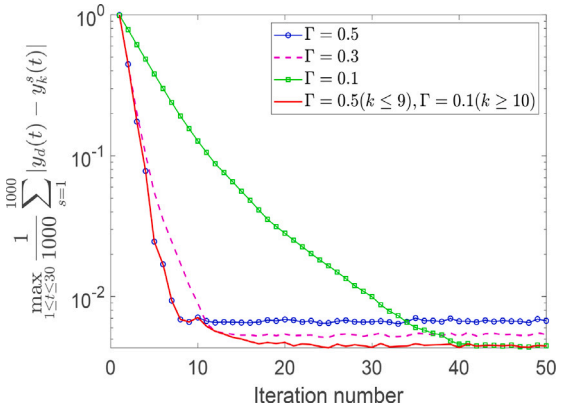

Fig. 8. Maximum mean output tracking errors for different learning gains under $\Phi = -0.5Cf$.

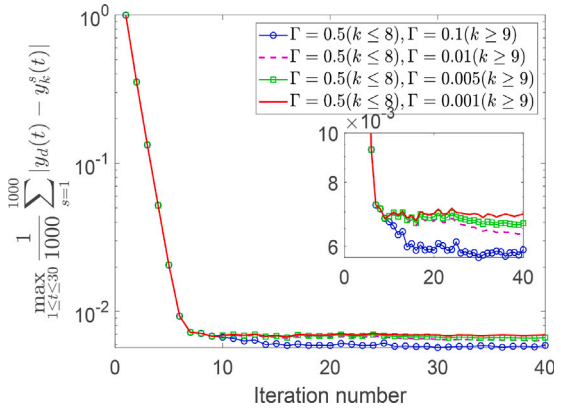

Fig. 9. Maximum mean output tracking errors for iteration-varying gain under $\Phi = -0.8Cf$.

Under spectral radius condition, by point-wise analysis technique, we first show that continuous nonlinearity can ensure bounded convergence. This surely extends the results presented in [31], which require locally Lipschitz nonlinearity with additional growth constraints for bounded convergence. Then, under spectral radius condition and the local Lipschitz nonlinearity, we establish the relations between steady-state tracking performance and state feedback as well as the standard deviations of measurement noises. Under fixed measurement accuracies of output and state, the findings reveal how the standard deviations shape the steady-state tracking performance. Furthermore, the relations also reveal how state feedback as well as the standard deviations shapes steady-state tracking performance, and how to utilize state feedback to enhance the steady-state tracking performance. Finally, under

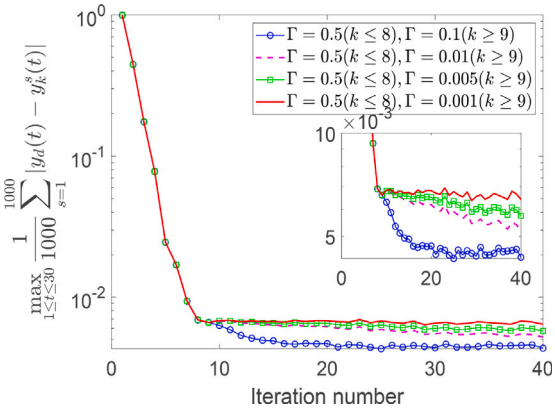

Fig. 10. Maximum mean output tracking errors for iteration-varying gain under $\Phi = -0.5Cf$.

spectral radius condition and the local Lipschitz nonlinearity, we establish the relations between transient tracking performance and state feedback as well as the standard deviations of measurement noises, which reveal how state feedback shapes the transient tracking performance and how to utilize state feedback to enhance the transient tracking performance.

CRediT authorship contribution statement

Yamiao Zhang: Writing – review & editing, Writing – original draft, Methodology, Investigation. **Jian Liu:** Writing – review & editing, Project administration, Methodology, Investigation. **Yuanshi Zheng:** Writing – review & editing, Methodology, Funding acquisition. **Xiaoe Ruan:** Validation, Methodology.

Declaration of competing interest

No conflict of interest exists in the submission of this manuscript, and the manuscript is approved by all authors for publication. I would like to declare on behalf of my co- authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62273267, the Key R&D program of Shaanxi Province under Grant 2025CY-YBXM-100 and the Natural Science Foundation of Shaanxi Province under Grant 2024JC-YBMS-008.

Appendix

Proof of Lemma 1. Since the nonnegative sequence ϕ_k is bounded, there exists the supremum of ϕ_k. This means C_ϕ is well defined. According to the definition of supremum, $|\phi_k| \le C_\phi$ for all $k \in \mathbb{N}$. By (7) we have

$$e_{k+1} \le \theta e_k + [1-\theta]\frac{C_\phi}{1-\theta} \text{ for all } k \in \mathbb{N}. \tag{77}$$

Clearly, $e_1 \le \max\left\{e_1, \frac{C_\phi}{1-\theta}\right\}$, which means (8) holds for $k = 1$.

Next, we move to prove (8) for $k = i+1$ provided that (8) holds for $k = i (i \ge 1)$. Using (77) and $\frac{C_\phi}{1-\theta} \le \max\left\{e_1, \frac{C_\phi}{1-\theta}\right\}$ together with the induction hypothesis $e_i \le \max\left\{e_1, \frac{C_\phi}{1-\theta}\right\}$ yields $e_{i+1} \le \max\left\{e_1, \frac{C_\phi}{1-\theta}\right\}$. It follows from mathematical induction that (8) holds for all $k \in \mathbb{N}$. This proves Lemma 1.

Data availability

No data was used for the research described in the article.

References

[1] Seel T, Schauer T, Raisch J. Monotonic convergence of iterative learning control systems with variable pass length. Internat J Control 2017;90(3):393–406.
[2] Yang Y. Injection molding control: from process to quality (Ph.D. thesis), Hong Kong University of Science and Technology; 2004.
[3] Arimoto S, Kawamura S, Miyazaki F. Bettering operation of robots by learnig. J Robot Syst 1984;1(2):123–40.
[4] Liu Y, Yao X, Zhao W. Distributed neural-based fault-tolerant control of multiple flexible manipulators with input saturations. Automatica 2023;156:111202.
[5] Jiang X, Chi M, Chen X, Yan H, Huang T. Tracking and regulation performance limitations of networked control systems over erasure channel with input quantization. IEEE Trans Autom Control 2022;67(9):4862–9.
[6] Jiang X, Yan H, Li B, Zheng S, Huang T, Shi P. Optimal tracking of networked control systems over the fading channel with packet loss. IEEE Trans Autom Control 2024;69(4):2652–9.
[7] Ahn H, Chen Y, Moore K. Iterative learning control: Brief survey and categorization. IEEE Trans Syst Man Cybern Part C 2007;37(6):1099–121.
[8] Bristow D, Tharayil M, Alleyne A. A survey of iterative learning control. IEEE Control Syst Mag 2006;26(3):96–114.
[9] Qiu W, Wang J, Shen D. Quantized iterative learning control for impulsive differential inclusion systems with data dropouts. ISA Trans 2024;145:285–97.
[10] Xu J. A survey on iterative learning control for nonlinear systems. Internat J Control 2011;84(7):1275–94.
[11] Zhao X, Wang Y. Improved point-to-point iterative learning control for batch processes with unknown batch-varying initial state. ISA Trans 2022;125(6):290–9.
[12] Huang D, Yang W, Huang T, Qin N, Chen Y, Tan Y. Iterative learning operation control of high-speed trains with adhesion dynamics. IEEE Trans Control Syst Technol 2021;29(6):2598–608.
[13] Cao X, Fečkan M, Shen D, Wang J. Iterative learning control for multi-agent systems with impulsive consensus tracking. Nonlinear Anal Model Control 2021;26(1):130–50.
[14] Shen D, Xu J. Distributed adaptive iterative learning control for nonlinear multiagent systems with state constraints. Internat J Adapt Control Signal Process 2017;31(12):1779–807.
[15] Blackwell M, Tutty O, Rogers E, Sandberg R. Iterative learning control applied to a nonlinear vortex panel model for improved aerodynamic load performance of wind turbines with smart rotors. Internat J Control 2016;89(1):55–68.
[16] Shen D. Iterative learning control with incomplete information: A survey. IEEE/CAA J Autom Sin 2018;5(5):885–901.
[17] Bien Z, Huh K. High-order iterative learning control algorithm. IEE Proc D Control Theory Appl 1989;136(3):105–12.
[18] Liu J, Zhang Y, Ruan X. Iterative learning control for a class of uncertain nonlinear systems with current state feedback. Int J Syst Sci 2019;50(10):1889–901.
[19] Ruan X, Bien Z, Wang Q. Convergence properties of iterative learning control processes in the sense of the lebesgue-p norm. Asian J Control 2012;14(4):1095–107.
[20] Xu J, Tan Y. Robust optimal design and convergence properties analysis of iterative learning control approaches. Automatica 2002;38(11):1867–80.
[21] Lee H, Bien Z. A note on convergence property of iterative learning controller with respect to sup norm. Automatica 1997;33(8):1591–3.
[22] Kurek J, Zaremba M. Iterative learning control synthesis based on 2-d system theory. IEEE Trans Autom Control 1993;38(1):121–5.
[23] Moore K, Chen Y, Vikas B. Monotonically convergent iterative learning control for linear discrete-time systems. Automatica 2005;41(9):1529–37.
[24] Park K, Bien Z. A study on iterative learning control with adjustment of learning interval for monotone convergence in the sense of supnorm. Asian J Control 2002;4(1):111–8.
[25] Park K, Bien Z. Intervalized iterative learning control for monotonic convergence in the sense of sup-norm. Internat J Control 2005;78(15):1218–27.
[26] Meng D, Jia Y, Du J, Yuan S. Feedback approach to design fast iterative learning controller for a class of time-delay systems. IET Control Theory Appl 2009;3(2):225–38.
[27] Liu J, Ruan X, Zheng Y. Iterative learning control for discrete-time systems with full learnability. IEEE Trans Neural Netw. Learn Syst 2022;33(2):629–43.
[28] Zhang Y, Liu J, Ruan X. Equivalence and convergence of two iterative learning control schemes with state feedback. Internat J Robust Nonlinear Control 2022;32(3):1561–82.
[29] Bu X, Hou Z, Jin S, Chi R. An iterative learning control design approach for networked control systems with data dropouts. Internat J Robust Nonlinear Control 2016;26:91–109.
[30] Liu J, Ruan X, Zhang Q. A joint control protocol for a class of uncertain nonlinear systems with iteration-varying trial length. Int J Syst Sci 2020;51(12):2276–92.
[31] Meng D, Moore K. Contraction mapping-based robust convergence of iterative learning control with uncertain, locally lipschitz nonlinearity. IEEE Trans Syst Man Cybern: Syst 2020;50(2):442–54.
[32] Wu Y, Meng D, Wang J, Cai K. Observer-based distributed methods for learning control systems. IEEE Trans Syst Man Cybern: Syst 2023;53(5):2983–94.
[33] Zhang J, Meng D. Generalized trackability analysis of learning control systems: A frequency-domain method. IEEE Trans Syst Man Cybern: Syst 2025;55(4):2806–20.
[34] Chen H. Almost sure convergence of iterative learning control for stochastic systems. Sci China 2003;46(1):67–79.
[35] Shen D, Wang Y. Iterative learning control for networked stochastic systems with random packet losses. Internat J Control 2015;88(5):959–68.
[36] Wang D. Convergence and robustness of discrete time nonlinear systems with iterative learning control. Automatica 1998;34(11):1445–8.