

SAC-enhanced iterative learning control with two-layer barrier mapping for nonlinear MIMO systems with asymmetric input constraints

Hongfeng Tao

taohongfeng@jiangnan.edu.cn

Jiangnan University

Fei Shao

Jiangnan University

Zhihe Zhuang

Jiangnan University

Wojciech Paszke

University of Zielona Góra

Research Article

Keywords:

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11319856/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

SAC-enhanced iterative learning control with two-layer barrier mapping for nonlinear MIMO systems with asymmetric input constraints

Hongfeng Tao^{1*}, Fei Shao¹, Zhihe Zhuang¹, Wojciech Paszke²

¹Key Laboratory of Advanced Process Control for Light Industry
(Ministry of Education), Jiangnan University, Wuxi, China.

²Institute of Automation, Electronic and Electrical Engineering,
University of Zielona Góra, Zielona Góra, Poland.

*Corresponding author(s). E-mail(s): taohongfeng@jiangnan.edu.cn;

Abstract

This paper considers tracking control of nonlinear multiple-input multiple-output (MIMO) repetitive systems subject to nonrepetitive disturbances and asymmetric input constraints. When iterative learning control (ILC) is augmented with residual compensation, feasibility must be maintained for both the stored learning input and the input applied to the plant. A proportional-derivative ILC (PD-ILC) scheme with soft actor-critic (SAC) residual compensation and two recentered barrier mappings is proposed. The first mapping constrains the learning input after each PD-ILC update. The second mapping enforces the input constraints after the bounded SAC residual is added to the learning input. A nonlinear finite-horizon input-output map is constructed, and an exact iteration-domain tracking-error recursion is derived using a trajectory-dependent averaged Jacobian. Under the stated assumptions, an explicit tracking-error bound is obtained in a common weighted norm, and uniform ultimate boundedness is established. Simulations on a coupled nonlinear mechanical system show improved tracking performance under nonrepetitive disturbances while maintaining input feasibility. Ablation studies distinguish the roles of the two mappings in constraining the stored learning input and the applied control input.

Keywords: Nonlinear MIMO systems, asymmetric input constraints, iterative learning control, soft actor-critic, two-layer barrier mapping

1 Introduction

Iterative learning control (ILC) improves tracking performance in repetitive tasks by updating the control input using information from previous iterations over a finite horizon. Its fundamental principles, convergence properties, and implementation issues have been reviewed in [1]. Representative applications include metal-strip regulation, robotic path learning, and power electronic systems [2–4].

Nonrepetitive effects in ILC can arise from variations in either the tracking task or the operating conditions. For uncertain nonlinear multiple-input multiple-output (MIMO) systems, [5] investigated ILC with nonrepetitive reference trajectories, while [6] developed a model-free formulation for second-order MIMO nonlinear systems under varying tracking tasks. A different challenge arises when the reference remains fixed but disturbances or uncertainties change between iterations. Sampled-data convergence analysis for nonlinear nonaffine systems [7] and finite-time observer-based disturbance compensation [8] have been developed to address such variations. Adaptive ILC has also been investigated for two-dimensional nonlinear systems with nonrepetitive uncertainties [9]. Robust optimization [10] and model-based predictive ILC [11] incorporate uncertainty into the learning design, whereas set-to-set ILC [12] considers tracking within admissible regions rather than a single prescribed trajectory. Input-constrained data-driven ILC has further been studied for nonlinear batch processes with nonrepetitive disturbances [13].

Another line of research focuses on constructing and adapting the learning input from available process information. Indirect ILC combines observer-based prediction, feedback, and iterative learning to address model uncertainty and input delay [14]. Iterative dynamic linearization [15] and data-driven learning for discrete-time nonlinear MIMO systems [16] provide alternative ways to construct learning controllers.

Machine-learning-based ILC for nonrepetitive time-varying systems [17] and adaptive model matching for nonlinear systems [18] further extend the available learning mechanisms.

Reinforcement learning (RL) provides an additional mechanism for adapting control behavior under uncertain operating conditions. Liu et al. [19] combined two-dimensional ILC with deep reinforcement learning compensation for nonrepetitive uncertain batch processes. Soleimani et al. [20] incorporated RL into the adjustment of the ILC forgetting factor, while Tao et al. [21] combined reinforcement learning compensation with a Koopman-based representation of nonlinear batch dynamics. Beyond ILC, soft actor-critic (SAC) has been applied to active disturbance rejection control for path following under wind and wave disturbances [22]. The role of RL differs across these formulations: it may adjust the learning mechanism or provide compensation alongside the nominal controller. In the latter case, the compensating action changes the command applied during each iteration, and its interaction with input constraints becomes part of the control design.

Input constraints have been addressed through several ILC formulations. Constrained ILC for linear systems [23] and robustness analysis of optimization-based constrained learning [24] established important results for input-limited repetitive control. Data-driven methods also incorporated input constraints into optimal and norm-optimal learning designs [25, 26], while constrained ILC for linear time-varying systems has also been experimentally evaluated [27]. Adaptive ILC has been investigated for nonlinear systems subject to state and input constraints [28], and input-constrained learning has been applied to flexible structures [29].

Saturation-aware ILC has received particular attention. Zhang and Meng [30] analyzed convergence for saturated ILC systems with locally Lipschitz nonlinearities, while Xu et al. [31] investigated input saturation through a composite-energy-function

approach. Sebastian et al. [32] studied feedback-based ILC with separate saturation stages for the learning feedforward signal and the combined plant command. Related developments include adaptive neural control for nonlinear MIMO systems with asymmetric saturation and dead zones [33], as well as alternating-projection ILC for nonuniform trial lengths [34]. Barrier-based techniques offer another approach to constraint handling, including control barrier functions for safety-critical systems [35], barrier Lyapunov functions for output-constrained nonlinear control [36], and barrier-based model predictive control [37].

In the nonlinear MIMO repetitive setting considered here, the PD-ILC update determines the stored learning sequence, while a state-dependent SAC residual modifies the plant command during each iteration. This gives rise to two distinct input-feasibility requirements: the stored learning sequence must remain feasible after the ILC update, and the residual-modified command must satisfy the input bounds before being applied to the plant. The proposed scheme addresses both requirements through recentered barrier mappings and explicitly incorporates the corresponding corrections into the nonlinear iteration-domain tracking-error analysis.

The main contributions of this work are as follows.

(1) A proportional-derivative ILC (PD-ILC) scheme with SAC residual compensation and two recentered barrier mappings is developed for nonlinear MIMO repetitive systems subject to nonrepetitive disturbances and asymmetric input constraints. The two mappings act after the PD-ILC update and after SAC residual compensation, respectively, thereby maintaining strict interior feasibility of both the stored learning sequence and the applied control input.

(2) For the proposed scheme, an exact iteration-domain tracking-error recursion is derived from the nonlinear finite-horizon input-output map, accounting for barrier corrections, SAC residual-action increments, and nonrepetitive disturbances. This

recursion is used to establish an explicit tracking-error bound and uniform ultimate boundedness of the tracking error under the stated assumptions.

Comparisons and ablation studies on a coupled nonlinear MIMO system evaluate the tracking improvement provided by SAC residual compensation and the constraint-handling roles of the two barrier mappings.

Table 1 summarizes the principal notation. Lowercase time-indexed quantities denote within-iteration signals, whereas uppercase quantities denote the corresponding finite-horizon stacked sequences when applicable. The remainder of this paper is organized as follows. Sect. 2 formulates the nonlinear MIMO repetitive system and the control objective. Sect. 3 presents the SAC-enhanced PD-ILC scheme with two-layer barrier mapping. Sect. 4 establishes the barrier-mapping properties and the tracking-error bound. Sect. 5 reports numerical simulations, and Sect. 6 concludes the paper.

Table 1 Principal notations

Symbol	Description	Symbol	Description
$\mathbf{x}_k(t)$	System state	$\mathbf{U}_k$	Stacked applied control input sequence
$\mathbf{u}_k(t)$	Applied control input	$\mathbf{U}_k^I$	ILC learning input sequence
$\mathbf{d}_k(t)$	Nonrepetitive disturbance	$\mathbf{U}_k^S$	Stacked SAC residual-action sequence
$\mathbf{y}_k(t), \mathbf{v}_k(t)$	Primary and auxiliary outputs	$\mathbf{C}_k$	Combined pre-barrier control sequence
$\mathbf{e}_k(t), \mathbf{e}_k^v(t)$	Primary and auxiliary tracking errors	$\hat{\mathbf{U}}_{k+1}$	Candidate PD-ILC learning input sequence
$\mathbf{u}_k^S(t)$	SAC residual action	$\mathbf{E}_k, \mathbf{E}_k^v$	Stacked tracking-error sequences
$\mathbf{s}_k(t)$	SAC state vector	$\mathbf{L}_p, \mathbf{L}_v$	PD-ILC learning matrices
$\tilde{\pi}_\theta$	SAC policy for normalized actions	$\mathcal{B}_1, \mathcal{B}_2$	First- and second-layer barrier mappings
β_k, ζ_k	First- and second-layer barrier corrections	$\mathbf{G}_k$	Averaged Jacobian of the input-output map
ω_k	Bounded perturbation term		

2 Problem Formulation

2.1 Nonlinear MIMO Repetitive System

To describe the coupled nonlinear dynamics, consider a discrete-time nonlinear MIMO repetitive system operating over a fixed horizon of T sampling steps, $t = 0, 1, \dots, T-1$:

$$\mathbf{x}_k(t+1) = f(\mathbf{x}_k(t), \mathbf{u}_k(t), \mathbf{d}_k(t)), \quad (1)$$

$$\mathbf{y}_k(t) = g(\mathbf{x}_k(t)), \quad (2)$$

where $k \in \mathbb{N}$ denotes the iteration index, $\mathbf{x}_k(t) \in \mathbb{R}^n$, $\mathbf{u}_k(t) \in \mathbb{R}^m$, and $\mathbf{d}_k(t) \in \mathbb{R}^{n_d}$. The output $\mathbf{y}_k(t) \in \mathbb{R}^p$ represents the primary controlled variable. The analysis permits $m \geq p$. A measurable auxiliary output $\mathbf{v}_k(t) = h(\mathbf{x}_k(t)) \in \mathbb{R}^p$ is introduced for the learning update.

All iterations share the same initial condition

$$\mathbf{x}_k(0) = \mathbf{x}_0, \quad \forall k, \quad \mathbf{y}_d(0) = g(\mathbf{x}_0), \quad (3)$$

which gives $\mathbf{e}_k(0) = 0$.

For the desired primary output trajectory $\mathbf{y}_d(t)$ and auxiliary trajectory $\mathbf{v}_d(t)$, the tracking errors are defined as

$$\mathbf{e}_k(t) = \mathbf{y}_d(t) - \mathbf{y}_k(t), \quad (4)$$

$$\mathbf{e}_k^v(t) = \mathbf{v}_d(t) - \mathbf{v}_k(t). \quad (5)$$

The auxiliary error is formed from the measurable output $\mathbf{v}_k$ and provides the derivative-type channel in the PD-ILC update. Its relation to the backward difference

of the primary tracking error is represented explicitly by the discrepancy term $\boldsymbol{\eta}_k$ in Sect. 4.

The physical input set is defined as

$$\mathcal{U} = \{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}_{\min} \preceq \mathbf{u} \preceq \mathbf{u}_{\max}\}, \quad (6)$$

where $\mathbf{u}_{\min}, \mathbf{u}_{\max} \in \mathbb{R}^m$ denote the lower and upper input bounds, respectively, and $\preceq$ represents the componentwise inequality.

For a prescribed margin $\varepsilon_u > 0$, define the tightened bounds $\boldsymbol{\ell} = \mathbf{u}_{\min} + \varepsilon_u \mathbf{1}_m$ and $\mathbf{r} = \mathbf{u}_{\max} - \varepsilon_u \mathbf{1}_m$, where $\mathbf{1}_m$ denotes the m -dimensional vector of ones, and $\boldsymbol{\ell}, \mathbf{r} \in \mathbb{R}^m$ are the tightened lower and upper input-bound vectors, respectively.

The resulting tightened feasible set and its strict interior are defined as

$$\mathcal{U}_{\varepsilon_u} = \{\mathbf{u} \in \mathbb{R}^m : \boldsymbol{\ell} \preceq \mathbf{u} \preceq \mathbf{r}\}, \quad \mathcal{U}_{\varepsilon_u}^{\circ} = \{\mathbf{u} \in \mathbb{R}^m : \boldsymbol{\ell} \prec \mathbf{u} \prec \mathbf{r}\}. \quad (7)$$

The interior of the tightened set is used in the barrier-mapping optimization, leaving a margin from the physical input bounds.

Define the stacked input and output sequences as

$$\begin{aligned} \mathbf{U}_k &= [\mathbf{u}_k^{\top}(0), \dots, \mathbf{u}_k^{\top}(T-1)]^{\top} \in \mathbb{R}^{mT}, \\ \mathbf{Y}_k &= [\mathbf{y}_k^{\top}(1), \dots, \mathbf{y}_k^{\top}(T)]^{\top} \in \mathbb{R}^{pT}. \end{aligned} \quad (8)$$

The reference sequences $\mathbf{Y}_d, \mathbf{V}_d$ and error sequences $\mathbf{E}_k, \mathbf{E}_k^v$ are formed analogously over $t = 1, \dots, T$ and belong to $\mathbb{R}^{pT}$. The implementation uses one-step input-output alignment: $\mathbf{u}_k(t)$ drives the state transition associated with $\mathbf{y}_k(t+1)$. The condition $\mathbf{e}_k(0) = 0$ in (3) supplies the boundary value for the first row of the backward-difference matrix in Sect. 4.

Define the corresponding closed sequence set and its strict interior as

$$\Omega_{\varepsilon_u} = (\mathcal{U}_{\varepsilon_u})^T, \quad \Omega_{\varepsilon_u}^\circ = (\mathcal{U}_{\varepsilon_u}^\circ)^T, \quad (9)$$

so that $\Omega_{\varepsilon_u}^\circ \subset \Omega_{\varepsilon_u} \subset \mathcal{U}^T$.

2.2 Finite-Horizon Nonlinear Input–Output Map

The disturbance sequence in iteration k is stacked as

$$\mathbf{d}_k = [\mathbf{d}_k^\top(0), \dots, \mathbf{d}_k^\top(T-1)]^\top \in \mathcal{D} \subset \mathbb{R}^{n_d T}, \quad (10)$$

where $\mathcal{D}$ is the set of admissible disturbance sequences. Let $\pi_t : \mathbb{R}^{n_d T} \rightarrow \mathbb{R}^{n_d}$ project onto the component at time t , and define the pointwise disturbance set

$$\mathbb{D} = \bigcup_{t=0}^{T-1} \pi_t(\mathcal{D}) \subset \mathbb{R}^{n_d}. \quad (11)$$

If $\mathcal{D}$ is compact, each projected set $\pi_t(\mathcal{D})$ is compact. The finite horizon then makes $\mathbb{D}$ compact. For an admissible input sequence $\mathbf{U} \in \Omega_{\varepsilon_u}$ and disturbance sequence $\mathbf{d} \in \mathcal{D}$, define the finite-horizon state trajectory recursively by

$$\begin{aligned} \mathbf{X}_0(\mathbf{U}, \mathbf{d}) &= \mathbf{x}_0, \\ \mathbf{X}_{t+1}(\mathbf{U}, \mathbf{d}) &= f(\mathbf{X}_t(\mathbf{U}, \mathbf{d}), \mathbf{u}(t), \mathbf{d}(t)), \quad t = 0, \dots, T-1, \end{aligned} \quad (12)$$

where $\mathbf{u}(t)$ and $\mathbf{d}(t)$ are the time- t blocks of $\mathbf{U}$ and $\mathbf{d}$. The corresponding primary-output component is

$$\mathcal{H}_t(\mathbf{U}, \mathbf{d}) = g(\mathbf{X}_t(\mathbf{U}, \mathbf{d})), \quad t = 1, \dots, T. \quad (13)$$

Thus, the nonlinear finite-horizon input–output map induced by the original state dynamics is

$$\mathcal{H}(\mathbf{U}, \mathbf{d}) = [\mathcal{H}_1^\top(\mathbf{U}, \mathbf{d}), \dots, \mathcal{H}_T^\top(\mathbf{U}, \mathbf{d})]^\top. \quad (14)$$

Hence, in iteration k , $\mathbf{Y}_k = \mathcal{H}(\mathbf{U}_k, \mathbf{d}_k)$ and $\mathbf{E}_k = \mathbf{Y}_d - \mathcal{H}(\mathbf{U}_k, \mathbf{d}_k)$.

2.3 Control Objective and Regularity Conditions

The objective is to maintain interior feasibility of both the ILC learning input and the applied control input at every iteration. The stacked tracking error is required to be uniformly ultimately bounded (UUB):

$$\mathbf{U}_k^I, \mathbf{U}_k \in \Omega_{\varepsilon_u}^\circ, \quad \forall k, \quad \limsup_{k \rightarrow \infty} \|\mathbf{E}_k\|_2 \leq \epsilon_E. \quad (15)$$

Here, $\mathbf{U}_k^I$ is the ILC learning input updated across iterations, and $\mathbf{U}_k$ is the applied control input.

Assumption 1 The transition $f(\mathbf{x}, \mathbf{u}, \mathbf{d})$ is continuous in $(\mathbf{x}, \mathbf{u}, \mathbf{d})$ and continuously differentiable with respect to $\mathbf{x}$ and $\mathbf{u}$. Its partial derivatives $\partial f / \partial \mathbf{x}$ and $\partial f / \partial \mathbf{u}$ are jointly continuous in $(\mathbf{x}, \mathbf{u}, \mathbf{d})$ on an open neighborhood containing the finite-horizon trajectories generated by $\mathbf{U} \in \Omega_{\varepsilon_u}$ and $\mathbf{d} \in \mathcal{D}$. The primary-output map $g(\mathbf{x})$ is continuously differentiable on the corresponding state region, the auxiliary-output map $h(\mathbf{x})$ is continuous there, and the admissible disturbance-sequence set $\mathcal{D} \subset \mathbb{R}^{n_d T}$ is compact.

Under Assumption 1 and the finite horizon $T < \infty$, repeated application of the chain rule to (12) shows that $\mathcal{H}(\mathbf{U}, \mathbf{d})$ is continuously differentiable with respect to $\mathbf{U}$. Its Jacobian $\mathbf{J}_{\mathcal{H}}(\mathbf{U}, \mathbf{d}) := \partial \mathcal{H}(\mathbf{U}, \mathbf{d}) / \partial \mathbf{U}$ is jointly continuous in $(\mathbf{U}, \mathbf{d})$ on an open neighborhood of $\Omega_{\varepsilon_u} \times \mathcal{D}$. Sect. 4 derives an explicit bound on ϵ_E using this regularity condition and the contraction condition stated there.

3 Controller Design

3.1 Controller Architecture

Figure 1 illustrates the transition from the completed k -th iteration to execution of iteration $k + 1$. After iteration k , PD-ILC generates the candidate learning sequence $\hat{U}_{k+1}$, which is mapped by $\mathcal{B}_1$ to U_{k+1}^I and stored in the iteration memory. During iteration $k + 1$, the stored sequence is read as $u_{k+1}^I(t)$, combined with the SAC residual $u_{k+1}^S(t)$, and mapped by $\mathcal{B}_2$ to the applied plant input $u_{k+1}(t)$.

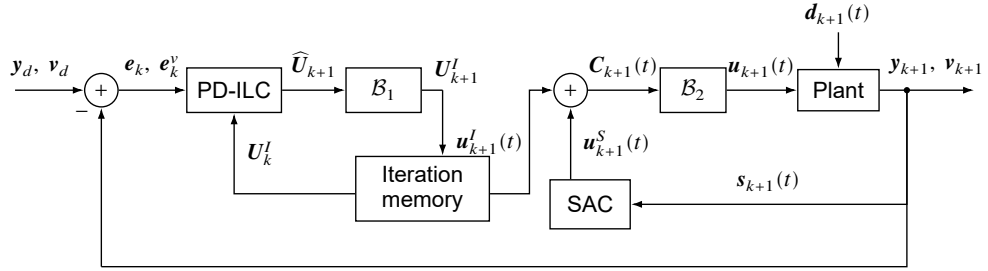


Fig. 1 Overall architecture of the proposed SAC-enhanced PD-ILC with two-layer barrier mapping. The SAC state is constructed from measurements, reference signals, and stored histories as specified in Sects. 3.4 and 5.2

3.2 Recentered Barrier Mapping

To maintain the learning sequence and applied control signal inside the tightened feasible set, define the stacked bounds

$$U_{\min} = \mathbf{1}_T \otimes \ell, \quad U_{\max} = \mathbf{1}_T \otimes r, \quad (16)$$

where $\mathbf{1}_T$ is the T -dimensional vector of ones and $U_{\min}, U_{\max} \in \mathbb{R}^{mT}$. The strict feasible sequence set in (9) can equivalently be written as

$$\Omega_{\varepsilon_u}^o = \{U \in \mathbb{R}^{mT} : U_{\min} \prec U \prec U_{\max}\}. \quad (17)$$

Assumption 2 The zero input sequence lies in the strict interior of the tightened feasible set, $0 \in \Omega_{\varepsilon_u}^\circ$.

On $\Omega_{\varepsilon_u}^\circ$, define the logarithmic barrier

$$b(\mathbf{U}) = -\mathbf{1}_{mT}^\top \ln(\mathbf{U} - \mathbf{U}_{\min}) - \mathbf{1}_{mT}^\top \ln(\mathbf{U}_{\max} - \mathbf{U}). \quad (18)$$

The logarithms act elementwise. For asymmetric bounds, the barrier gradient at the origin is generally nonzero.

Gradient recentering [37] ensures $\nabla \Phi(0) = 0$. Recentering preserves the Hessian of $b(\mathbf{U})$, and $\Phi(\mathbf{U})$ tends to infinity as $\mathbf{U}$ approaches the boundary of the feasible set:

$$\Phi(\mathbf{U}) = b(\mathbf{U}) - b(0) - \nabla b(0)^\top \mathbf{U}. \quad (19)$$

$$\begin{aligned} \Phi(0) &= 0, & \nabla \Phi(0) &= 0, & \nabla^2 \Phi(\mathbf{U}) &= \nabla^2 b(\mathbf{U}), \\ \Phi(\mathbf{U}) &\geq 0, & \mathbf{U} &\in \Omega_{\varepsilon_u}^\circ. \end{aligned} \quad (20)$$

The last inequality follows from the strict convexity of b and the supporting-hyperplane inequality at the interior point $\mathbf{U} = 0$.

For $\mathbf{Z} \in \mathbb{R}^{mT}$ and $\kappa > 0$, define the recentered barrier mapping

$$\mathcal{B}_\kappa(\mathbf{Z}) = \arg \min_{\mathbf{U} \in \Omega_{\varepsilon_u}^\circ} \left\{ \frac{1}{2} \|\mathbf{U} - \mathbf{Z}\|_2^2 + \kappa \Phi(\mathbf{U}) \right\}. \quad (21)$$

The quadratic term penalizes deviations from the candidate input $\mathbf{Z}$, whereas the barrier term penalizes approach to the tightened feasible-set boundary. The parameter κ sets their relative weight. The two control layers use $\mathcal{B}_1 = \mathcal{B}_{\kappa_1}$ and $\mathcal{B}_2 = \mathcal{B}_{\kappa_2}$.

Both terms in (21) are separable across time samples and input channels. For channel i , let $\ell_i < u < r_i$ and define

$$\begin{aligned} b_i(u) &= -\ln(u - \ell_i) - \ln(r_i - u), \\ \phi_i(u) &= b_i(u) - b_i(0) - b'_i(0)u, \\ b'_i(u) &= -\frac{1}{u - \ell_i} + \frac{1}{r_i - u}. \end{aligned} \tag{22}$$

The corresponding scalar map is

$$\mathcal{B}_{\kappa,i}(z) = \arg \min_{\ell_i < u < r_i} \left\{ \frac{1}{2}(u - z)^2 + \kappa \phi_i(u) \right\}. \tag{23}$$

Its minimizer satisfies the scalar first-order condition

$$F_{\kappa,i}(u; z) := u - z + \kappa [b'_i(u) - b'_i(0)] = 0. \tag{24}$$

The scalar objective is strictly convex and diverges at both interval endpoints, so this equation has a unique interior root for every $z \in \mathbb{R}$. In the numerical implementation, the root is evaluated using safeguarded Newton steps with midpoint fallback, a bracket whose endpoints lie strictly inside (ℓ_i, r_i) , a tolerance of $2 \times 10^{-10}(1 + |z|)$, and at most 80 iterations. Iterates remain strictly inside (ℓ_i, r_i) . By separability, $\mathcal{B}_\kappa$ is evaluated componentwise over all time samples and input channels.

3.3 PD-ILC Update and First-Layer Barrier Mapping

The PD-ILC learning update is

$$\hat{\mathbf{U}}_{k+1} = \mathbf{U}_k^I + \mathbf{L}_p \mathbf{E}_k + \mathbf{L}_v \mathbf{E}_k^v, \tag{25}$$

$$\mathbf{U}_{k+1}^I = \mathcal{B}_1(\hat{\mathbf{U}}_{k+1}). \tag{26}$$

Equivalently, the candidate is updated componentwise after completing iteration k :

$$\hat{\mathbf{u}}_{k+1}(t) = \mathbf{u}_k^I(t) + \mathbf{K}_p \mathbf{e}_k(t+1) + \mathbf{K}_v \mathbf{e}_k^v(t+1), \quad t = 0, \dots, T-1.$$

Here,

$$\mathbf{L}_p = \mathbf{I}_T \otimes \mathbf{K}_p, \quad \mathbf{L}_v = \mathbf{I}_T \otimes \mathbf{K}_v.$$

The matrices $\mathbf{K}_p, \mathbf{K}_v \in \mathbb{R}^{m \times p}$ are the PD-ILC learning gains for the tracking and auxiliary-output errors, respectively.

Componentwise, $\mathcal{B}_1$ uses the scalar mapping in (23)–(24) with $\kappa = \kappa_1$. Define the first-layer correction as

$$\beta_k = \mathbf{U}_{k+1}^I - \hat{\mathbf{U}}_{k+1}. \quad (27)$$

Proposition 4.1 ensures $\mathbf{U}_{k+1}^I \in \Omega_{\varepsilon_u}^\circ$ for every candidate $\hat{\mathbf{U}}_{k+1}$.

3.4 SAC Residual Compensator

Soft actor–critic (SAC) uses maximum-entropy reinforcement learning to promote exploration in continuous-action policies [38]. The SAC policy generates a bounded residual action through the squashed action parametrization. The SAC state vector is constructed using information available before action selection at time t :

$$\begin{aligned} \mathbf{s}_k(t) = & \left[(\mathbf{s}_k^{\text{core}}(t))^\top, (\mathbf{s}_k^{\text{hist}}(t))^\top, (\mathbf{s}^{\text{time}}(t))^\top, \right. \\ & \left. (\mathbf{s}_k^{\text{margin}}(t))^\top, (\mathbf{s}_k^{\text{cross}}(t))^\top \right]^\top. \end{aligned} \quad (28)$$

The state vector includes current measurements, past data, time information, distances to the input bounds, and data from the preceding iteration. The disturbance realization is not supplied to the policy. The implementation-specific contents and dimensions are given in Sect. 5.2.

The reparameterized policy generates a normalized squashed action $\mathbf{b}_k(t)$ and the physical residual action $\mathbf{a}_k(t)$:

$$\begin{aligned} \mathbf{z}_k(t) &= \boldsymbol{\mu}_\theta(\mathbf{s}_k(t)) + \boldsymbol{\sigma}_\theta(\mathbf{s}_k(t)) \odot \boldsymbol{\xi}, & \boldsymbol{\xi} &\sim \mathcal{N}(0, \mathbf{I}), \\ \mathbf{b}_k(t) &= \tanh \mathbf{z}_k(t), \\ \mathbf{a}_k(t) &= \mathbf{u}_k^S(t) = \bar{\mathbf{a}} \odot \mathbf{b}_k(t). \end{aligned} \tag{29}$$

Here, $\bar{\mathbf{a}} \in \mathbb{R}_{>0}^m$ is the prescribed componentwise amplitude bound for the SAC action. Hence, $|u_{k,i}^S(t)| < \bar{a}_i$ for $i = 1, \dots, m$. Since the barrier mapping is separable across times and channels, let $\mathcal{B}_{2,t}$ denote its second-layer component at time t . Define

$$\begin{aligned} \mathbf{C}_k(t) &= \mathbf{u}_k^I(t) + \mathbf{u}_k^S(t), \\ \mathbf{u}_k(t) &= \mathcal{B}_{2,t}(\mathbf{C}_k(t)), \\ \boldsymbol{\zeta}_k(t) &= \mathbf{u}_k(t) - \mathbf{C}_k(t). \end{aligned} \tag{30}$$

Overbarred variables in the reward represent quantities normalized by fixed scales. Throughout, δ denotes time-domain increments and Δ denotes iteration-domain increments, with endpoints specified below. Define the backward time-domain SAC action increment and the backward difference between completed iterations as

$$\begin{aligned} \delta \mathbf{u}_k^S(t) &= \mathbf{u}_k^S(t) - \mathbf{u}_k^S(t-1), \\ \nabla_k \bar{\mathbf{e}}_k(t+1) &= \bar{\mathbf{e}}_k(t+1) - \bar{\mathbf{e}}_{k-1}(t+1), & k &\geq 1. \end{aligned}$$

Here, $\mathbf{u}_k^S(-1) = 0$ at the first time sample of each iteration, and $\delta \bar{\mathbf{u}}_k^S(t)$ denotes the normalized action increment. Set $\nabla_k \bar{\mathbf{e}}_0(t+1) = 0$. The reward uses a backward difference between completed iterations, distinct from the forward iteration increment $\Delta \mathbf{X}_k = \mathbf{X}_{k+1} - \mathbf{X}_k$ used in the analysis.

Collect the reward-related quantities and define their weighting matrix as

$$\begin{aligned}\boldsymbol{\psi}_k(t) &= \left[\bar{\mathbf{e}}_k^\top(t+1), (\bar{\mathbf{e}}_k^v(t+1))^\top, (\bar{\mathbf{u}}_k^S(t))^\top, \right. \\ &\quad \left. (\delta \bar{\mathbf{u}}_k^S(t))^\top, \bar{\boldsymbol{\zeta}}_k^\top(t), (\nabla_k \bar{\mathbf{e}}_k(t+1))^\top \right]^\top, \\ \mathbf{Q}_r &= \text{blkdiag}(q_e \mathbf{I}_p, q_v \mathbf{I}_p, q_u \mathbf{I}_m, q_s \mathbf{I}_m, q_b \mathbf{I}_m, q_c \mathbf{I}_p).\end{aligned}$$

The per-step reward is

$$r_k(t) = -\|\boldsymbol{\psi}_k(t)\|_{\mathbf{Q}_r}^2. \quad (31)$$

Here, $\|\boldsymbol{\psi}\|_{\mathbf{Q}_r}^2 = \boldsymbol{\psi}^\top \mathbf{Q}_r \boldsymbol{\psi}$. The SAC action bounds, reward weights and normalization scales, and network and training parameters are given in Sect. 5.2.

The policy log density is defined on the normalized action $\mathbf{b} = \tanh \mathbf{z}$, with $\mathbf{z} = \boldsymbol{\mu}_\theta(\mathbf{s}) + \boldsymbol{\sigma}_\theta(\mathbf{s}) \odot \boldsymbol{\xi}$:

$$\log \tilde{\pi}_\theta(\mathbf{b} \mid \mathbf{s}) = \log \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_\theta(\mathbf{s}), \text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{s}))) - \sum_{i=1}^m \log(1 - b_i^2). \quad (32)$$

For numerical stability, the logarithmic Jacobian terms are evaluated as $\log(1 - b_i^2 + 10^{-6})$. Let $\bar{Q}_j$ denote the target critics, $\chi_{\text{term}} \in \{0, 1\}$ indicate the terminal transition at the end of a T -step iteration, and $\mathbf{a} = \bar{\mathbf{a}} \odot \mathbf{b}$ be the physical action supplied to the critics. For the next state, $\mathbf{b}' \sim \tilde{\pi}_\theta(\cdot \mid \mathbf{s}')$ and $\mathbf{a}' = \bar{\mathbf{a}} \odot \mathbf{b}'$. The temperature objective follows the automatic entropy adjustment developed in [39]. The critic target and optimization objectives are

$$y_Q = r + \gamma(1 - \chi_{\text{term}}) \left[\min_{j=1,2} \bar{Q}_j(\mathbf{s}', \mathbf{a}') - \alpha \log \tilde{\pi}_\theta(\mathbf{b}' \mid \mathbf{s}') \right], \quad (33)$$

$$J_Q = \sum_{j=1}^2 \mathbb{E}[\text{Huber}(Q_j(\mathbf{s}, \mathbf{a}) - y_Q)], \quad (34)$$

$$J_\pi = \mathbb{E} \left[\alpha \log \tilde{\pi}_\theta(\mathbf{b} \mid \mathbf{s}) - \min_{j=1,2} Q_j(\mathbf{s}, \mathbf{a}) \right], \quad (35)$$

$$J_{\lambda_\alpha} = -\mathbb{E}[\lambda_\alpha \text{stopgrad}(\log \tilde{\pi}_\theta(\mathbf{b} \mid \mathbf{s}) + \mathcal{H}_0)], \quad \alpha = e^{\lambda_\alpha}. \quad (36)$$

Here, $\mathcal{H}_0$ denotes the target entropy, and stopgrad treats its argument as constant during differentiation. The critics receive the physical action $\mathbf{a}$, whereas the entropy terms in Eqs. (33), (35), and (36) use the log density $\log \tilde{\pi}_\theta(\mathbf{b} \mid \mathbf{s})$ in normalized action coordinates. Target networks are soft-updated at rate τ_{soft} . Training samples stochastic actions, whereas evaluation uses the deterministic action generated by a policy with fixed parameters, $\mathbf{u}_k^S(t) = \bar{\mathbf{a}} \odot \tanh \boldsymbol{\mu}_\theta(\mathbf{s}_k(t))$.

3.5 Second-Layer Barrier Mapping

The stacked sequences from Sect. 3.4 and their relations to the applied input are

$$\begin{aligned} \mathbf{U}_k^S &= [(\mathbf{u}_k^S(0))^\top, \dots, (\mathbf{u}_k^S(T-1))^\top]^\top, \\ \mathbf{C}_k &= [\mathbf{C}_k^\top(0), \dots, \mathbf{C}_k^\top(T-1)]^\top, \\ \boldsymbol{\zeta}_k &= [\boldsymbol{\zeta}_k^\top(0), \dots, \boldsymbol{\zeta}_k^\top(T-1)]^\top, \\ \mathbf{C}_k &= \mathbf{U}_k^I + \mathbf{U}_k^S, \end{aligned} \quad (37)$$

$$\mathbf{U}_k = \mathcal{B}_2(\mathbf{C}_k), \quad (38)$$

$$\boldsymbol{\zeta}_k = \mathbf{U}_k - \mathbf{C}_k. \quad (39)$$

Applying Eqs. (23) and (24) componentwise with $\kappa = \kappa_2$ gives $\mathbf{U}_k \in \Omega_{\varepsilon_u}^\circ$ for any $\mathbf{C}_k \in \mathbb{R}^{mT}$; $\boldsymbol{\zeta}_k$ records the second-layer correction.

Algorithm 1 summarizes the training and control procedure.

Algorithm 1 Two-Layer Barrier PD-ILC with SAC

```

1: Initialize  $\mathcal{B}_1, \mathcal{B}_2$ , SAC networks, and feasible  $\mathbf{U}_0^I$ 
2: Phase I: SAC training
3: for  $j = 1, 2$  do
4:   Initialize  $\mathbf{U}_0^I$  and historical information
5:   for  $k = 0, \dots, K_{\text{tr}}^{(j)} - 1$  do
6:     for  $t = 0, \dots, T - 1$  do
7:       Generate  $\mathbf{u}_k^S(t)$  from  $\mathbf{s}_k(t)$ , form  $\mathbf{C}_k(t)$ , and apply  $\mathbf{u}_k(t) = \mathcal{B}_{2,t}(\mathbf{C}_k(t))$ 
8:       Observe  $\mathbf{x}_k(t+1)$ ; compute  $\mathbf{e}_k(t+1)$ ,  $\mathbf{e}_k^v(t+1)$ , and reward
9:       Construct the next SAC state vector  $\mathbf{s}_k(t+1)$ 
10:      Set  $\chi_{\text{term}} = \mathbf{1}\{t = T - 1\}$ ; store  $(\mathbf{s}_k(t), \mathbf{u}_k^S(t), r_k(t), \mathbf{s}_k(t+1), \chi_{\text{term}})$ 
11:    end for
12:    Compute  $\mathbf{E}_k, \mathbf{E}_k^v$  and update  $\mathbf{U}_{k+1}^I = \mathcal{B}_1(\mathbf{U}_k^I + \mathbf{L}_p \mathbf{E}_k + \mathbf{L}_v \mathbf{E}_k^v)$ 
13:    Update the SAC networks from the replay buffer after the iteration
14:  end for
15: end for
16: Freeze the validated SAC policy and reinitialize feasible  $\mathbf{U}_0^I$ 
17: Phase II: iterative control
18: for  $k = 0, \dots, K_{\text{eval}} - 1$  do
19:   for  $t = 0, \dots, T - 1$  do
20:     Generate  $\mathbf{u}_k^S(t)$  from the frozen policy, form  $\mathbf{C}_k(t)$ , and apply  $\mathbf{u}_k(t) = \mathcal{B}_{2,t}(\mathbf{C}_k(t))$ 
21:   end for
22:   Compute  $\mathbf{E}_k, \mathbf{E}_k^v$  and update  $\mathbf{U}_{k+1}^I = \mathcal{B}_1(\mathbf{U}_k^I + \mathbf{L}_p \mathbf{E}_k + \mathbf{L}_v \mathbf{E}_k^v)$ 
23: end for

```

4 Theoretical Analysis

Proposition 4.1 *For every $\mathbf{Z} \in \mathbb{R}^{mT}$ and $\kappa > 0$, the mapping $\mathcal{B}_\kappa(\mathbf{Z})$ is uniquely defined and satisfies $\mathcal{B}_\kappa(\mathbf{Z}) \in \Omega_{\varepsilon_u}^\circ$. Moreover, $\mathcal{B}_\kappa(0) = 0$, the map is continuously differentiable with respect to $\mathbf{Z}$, and it is nonexpansive in the Euclidean norm.*

Proof The Hessian of (18) is

$$\begin{aligned} \nabla^2 b(\mathbf{U}) &= \text{diag}[(\mathbf{U} - \mathbf{U}_{\min})^{\odot(-2)} \\ &\quad + (\mathbf{U}_{\max} - \mathbf{U})^{\odot(-2)}] \succ 0, \end{aligned} \tag{40}$$

where the powers are taken elementwise. Since $\nabla^2 \Phi(\mathbf{U}) = \nabla^2 b(\mathbf{U})$, the objective in (21) has Hessian $\mathbf{I} + \kappa \nabla^2 b(\mathbf{U}) \succ 0$ and is strictly convex. Also, $b(\mathbf{U}) \rightarrow +\infty$ whenever $\mathbf{U}$ approaches the boundary of $\Omega_{\varepsilon_u}^\circ$. Thus, for each $\mathbf{Z}$ there is a unique interior minimizer $\mathbf{U} = \mathcal{B}_\kappa(\mathbf{Z})$.

The unique minimizer satisfies the first-order optimality condition

$$\mathbf{U} - \mathbf{Z} + \kappa[\nabla b(\mathbf{U}) - \nabla b(0)] = 0. \tag{41}$$

At $\mathbf{Z} = 0$, the objective equals $\frac{1}{2}\|\mathbf{U}\|_2^2 + \kappa\Phi(\mathbf{U}) \geq 0$, with equality at $\mathbf{U} = 0$. Uniqueness therefore gives $\mathcal{B}_\kappa(0) = 0$.

The derivative of the left side of (41) with respect to $\mathbf{U}$ is $\mathbf{I} + \kappa\nabla^2 b(\mathbf{U}) \succ 0$. By the implicit-function theorem,

$$D\mathcal{B}_\kappa(\mathbf{Z}) = [\mathbf{I} + \kappa\nabla^2 b(\mathbf{U})]^{-1}, \quad \mathbf{U} = \mathcal{B}_\kappa(\mathbf{Z}). \quad (42)$$

It follows that $0 \prec D\mathcal{B}_\kappa(\mathbf{Z}) \preceq \mathbf{I}$ and $\|D\mathcal{B}_\kappa(\mathbf{Z})\|_2 \leq 1$. Integrating the derivative along the segment between arbitrary $\mathbf{Z}_1$ and $\mathbf{Z}_2$ yields

$$\|\mathcal{B}_\kappa(\mathbf{Z}_1) - \mathcal{B}_\kappa(\mathbf{Z}_2)\|_2 \leq \|\mathbf{Z}_1 - \mathbf{Z}_2\|_2. \quad \square \quad (43)$$

Theorem 4.1 *Suppose that $\mathbf{U}_0^I \in \Omega_{\varepsilon_u}^\circ$ and $\kappa_1, \kappa_2 > 0$. For every iteration $k \geq 0$, the controller (25)–(38) satisfies*

$$\mathbf{U}_k^I \in \Omega_{\varepsilon_u}^\circ, \quad \mathbf{U}_k \in \Omega_{\varepsilon_u}^\circ \subset \mathcal{U}^T. \quad (44)$$

Proof Proposition 4.1 gives $\mathbf{U}_{k+1}^I = \mathcal{B}_1(\widehat{\mathbf{U}}_{k+1}) \in \Omega_{\varepsilon_u}^\circ$ for every candidate $\widehat{\mathbf{U}}_{k+1}$. It also gives $\mathbf{U}_k = \mathcal{B}_2(\mathbf{C}_k) \in \Omega_{\varepsilon_u}^\circ$ for every combined signal $\mathbf{C}_k$. The initial condition on $\mathbf{U}_0^I$ and induction establish both claims for all $k \geq 0$. The inclusion in (9) then gives applied-input feasibility. $\square$

Here, $\Delta\mathbf{X}_k := \mathbf{X}_{k+1} - \mathbf{X}_k$ for stacked sequences $\mathbf{X} = \mathbf{U}, \mathbf{Y}, \mathbf{E}, \mathbf{U}^I, \mathbf{U}^S, \boldsymbol{\zeta}$. In particular, $\Delta\mathbf{E}_k$ stacks $\Delta\mathbf{e}_k(t) := \mathbf{e}_{k+1}(t) - \mathbf{e}_k(t)$ for $t = 1, \dots, T$.

For any admissible $(\mathbf{U}, \mathbf{d})$, define the local Jacobians along the trajectory in (12) by

$$\begin{aligned} \mathbf{F}_{x,t} &:= \frac{\partial f}{\partial \mathbf{x}}(\mathbf{X}_t, \mathbf{u}(t), \mathbf{d}(t)), \\ \mathbf{F}_{u,t} &:= \frac{\partial f}{\partial \mathbf{u}}(\mathbf{X}_t, \mathbf{u}(t), \mathbf{d}(t)), \\ \mathbf{G}_{x,t} &:= \frac{\partial g}{\partial \mathbf{x}}(\mathbf{X}_t). \end{aligned} \quad (45)$$

Their arguments $(\mathbf{U}, \mathbf{d})$ are suppressed below. The chain rule gives

$$\frac{\partial \mathcal{H}_t}{\partial \mathbf{u}(\tau)} = \begin{cases} \mathbf{G}_{x,t} \mathbf{F}_{x,t-1} \cdots \mathbf{F}_{x,\tau+1} \mathbf{F}_{u,\tau}, & \tau < t, \\ 0, & \tau \geq t \end{cases}. \quad (46)$$

For $t = \tau + 1$, the product of $\mathbf{F}_{x,j}$ factors is empty and equals the identity. By causality, $\mathbf{J}_{\mathcal{H}}(\mathbf{U}, \mathbf{d})$ is block lower triangular; its entries depend on the state, input, and disturbance trajectory.

Define the output difference caused by disturbance variation as

$$\Delta_k = \mathcal{H}(\mathbf{U}_{k+1}, \mathbf{d}_{k+1}) - \mathcal{H}(\mathbf{U}_{k+1}, \mathbf{d}_k), \quad (47)$$

and let $\Delta \mathbf{U}_k = \mathbf{U}_{k+1} - \mathbf{U}_k$. The averaged finite-horizon sensitivity is

$$\mathbf{G}_k = \int_0^1 \mathbf{J}_{\mathcal{H}}(\mathbf{U}_k + s \Delta \mathbf{U}_k, \mathbf{d}_k) \, ds. \quad (48)$$

Lemma 4.1 *Under the regularity conditions in Sect. 2.3,*

$$\Delta \mathbf{Y}_k = \mathbf{G}_k \Delta \mathbf{U}_k + \Delta_k. \quad (49)$$

Proof Theorem 4.1 and convexity of $\Omega_{\varepsilon_u}^\circ$ ensure $\mathbf{U}_k + s \Delta \mathbf{U}_k \in \Omega_{\varepsilon_u}^\circ$ for $s \in [0, 1]$. With $\mathbf{d}_k$ fixed, the fundamental theorem of calculus gives $\mathcal{H}(\mathbf{U}_{k+1}, \mathbf{d}_k) - \mathcal{H}(\mathbf{U}_k, \mathbf{d}_k) = \mathbf{G}_k \Delta \mathbf{U}_k$. Adding and subtracting $\mathcal{H}(\mathbf{U}_{k+1}, \mathbf{d}_k)$ in $\mathbf{Y}_{k+1} - \mathbf{Y}_k$ yields (49). $\square$

Let $T_s > 0$ denote the sampling interval, and let $\mathbf{D} \in \mathbb{R}^{T \times T}$ be the backward-difference matrix with unit diagonal and -1 on its first subdiagonal. With $\mathbf{e}_k(0) = 0$, the vector $\delta \mathbf{E}_k = (\mathbf{D} \otimes \mathbf{I}_p) \mathbf{E}_k$ stacks the time-domain differences $\mathbf{e}_k(t) - \mathbf{e}_k(t-1)$.

Define

$$\boldsymbol{\eta}_k := \boldsymbol{E}_k^v - \frac{1}{T_s}(\boldsymbol{D} \otimes \boldsymbol{I}_p)\boldsymbol{E}_k, \quad (50)$$

$$\boldsymbol{L} = \boldsymbol{L}_p + \frac{1}{T_s}\boldsymbol{L}_v(\boldsymbol{D} \otimes \boldsymbol{I}_p). \quad (51)$$

Using (25), (27), and these definitions gives

$$\Delta \boldsymbol{U}_k^I = \boldsymbol{L}\boldsymbol{E}_k + \boldsymbol{L}_v\boldsymbol{\eta}_k + \boldsymbol{\beta}_k. \quad (52)$$

Taking iteration differences in (37)–(39) therefore yields

$$\Delta \boldsymbol{U}_k = \boldsymbol{L}\boldsymbol{E}_k + \boldsymbol{L}_v\boldsymbol{\eta}_k + \boldsymbol{\beta}_k + \Delta \boldsymbol{U}_k^S + \Delta \boldsymbol{\zeta}_k. \quad (53)$$

Since the desired trajectory is identical across iterations, Lemma 4.1 gives $\Delta \boldsymbol{E}_k = -\boldsymbol{G}_k \Delta \boldsymbol{U}_k - \boldsymbol{\Delta}_k$. Substitution into $\boldsymbol{E}_{k+1} = \boldsymbol{E}_k + \Delta \boldsymbol{E}_k$ establishes the exact error recursion

$$\boldsymbol{E}_{k+1} = (\boldsymbol{I} - \boldsymbol{G}_k \boldsymbol{L})\boldsymbol{E}_k + \boldsymbol{\omega}_k, \quad (54)$$

where

$$\boldsymbol{\omega}_k = -\boldsymbol{G}_k(\boldsymbol{L}_v\boldsymbol{\eta}_k + \boldsymbol{\beta}_k + \Delta \boldsymbol{U}_k^S + \Delta \boldsymbol{\zeta}_k) - \boldsymbol{\Delta}_k. \quad (55)$$

Define the auxiliary-output sequence by

$$\mathcal{V}(\boldsymbol{U}, \boldsymbol{d}) := [h^\top(\boldsymbol{X}_1(\boldsymbol{U}, \boldsymbol{d})), \dots, h^\top(\boldsymbol{X}_T(\boldsymbol{U}, \boldsymbol{d}))]^\top. \quad (56)$$

For fixed $\mathbf{x}_0$ and finite T , the reachable state sets are continuous images of the compact admissible set $\Omega_{\varepsilon_u} \times \mathcal{D}$. Assumption 1 therefore gives finite uniform bounds

$$\begin{aligned}\bar{E} &:= \sup_{\mathbf{U} \in \Omega_{\varepsilon_u}, \mathbf{d} \in \mathcal{D}} \|\mathbf{Y}_d - \mathcal{H}(\mathbf{U}, \mathbf{d})\|_2, \\ \bar{E}_v &:= \sup_{\mathbf{U} \in \Omega_{\varepsilon_u}, \mathbf{d} \in \mathcal{D}} \|\mathbf{V}_d - \mathcal{V}(\mathbf{U}, \mathbf{d})\|_2.\end{aligned}\tag{57}$$

Let $\mathbf{r}_u = \max\{|\ell|, |\mathbf{r}|\}$ componentwise, $R_U = \sqrt{T}\|\mathbf{r}_u\|_2$, and $R_S = \sqrt{T}\|\bar{\mathbf{a}}\|_2$. The input bounds imply $\|\mathbf{U}_k^I\|_2, \|\mathbf{U}_k\|_2 \leq R_U$ and $\|\mathbf{U}_k^S\|_2 \leq R_S$. Consequently,

$$\begin{aligned}\|\boldsymbol{\eta}_k\|_2 &\leq \bar{\eta} := \bar{E}_v + T_s^{-1}\|\mathbf{D} \otimes \mathbf{I}_p\|_2 \bar{E}, \\ \|\boldsymbol{\beta}_k\|_2 &\leq \bar{\beta} := 2R_U + \|\mathbf{L}_p\|_2 \bar{E} + \|\mathbf{L}_v\|_2 \bar{E}_v, \\ \|\boldsymbol{\zeta}_k\|_2 &\leq \bar{\zeta} := 2R_U + R_S.\end{aligned}\tag{58}$$

Continuity on the same compact domain also yields

$$\begin{aligned}\bar{G} &:= \sup_{\mathbf{U} \in \Omega_{\varepsilon_u}, \mathbf{d} \in \mathcal{D}} \|\mathbf{J}_{\mathcal{H}}(\mathbf{U}, \mathbf{d})\|_2 < \infty, \\ \bar{\Delta} &:= \sup_{\mathbf{U} \in \Omega_{\varepsilon_u}, \mathbf{d}, \mathbf{d}' \in \mathcal{D}} \|\mathcal{H}(\mathbf{U}, \mathbf{d}') - \mathcal{H}(\mathbf{U}, \mathbf{d})\|_2 < \infty.\end{aligned}\tag{59}$$

Using $\|\mathbf{G}_k\|_2 \leq \bar{G}$, $\|\Delta \mathbf{U}_k^S\|_2 \leq 2R_S$, and $\|\Delta \boldsymbol{\zeta}_k\|_2 \leq 2\bar{\zeta}$ in (55) gives

$$\|\boldsymbol{\omega}_k\|_2 \leq \bar{\omega} := \bar{G}(\|\mathbf{L}_v\|_2 \bar{\eta} + \bar{\beta} + 2R_S + 2\bar{\zeta}) + \bar{\Delta} < \infty.\tag{60}$$

Define the one-step input-output sensitivity and the diagonal block of the stacked learning matrix by

$$\mathbf{S}_t(\mathbf{U}, \mathbf{d}) := \frac{\partial \mathcal{H}_t(\mathbf{U}, \mathbf{d})}{\partial \mathbf{u}(t-1)} = \mathbf{G}_{x,t} \mathbf{F}_{u,t-1}, \quad \mathbf{K}_0 := \mathbf{K}_p + T_s^{-1} \mathbf{K}_v, \quad t = 1, \dots, T.\tag{61}$$

Assumption 3 There is a constant $r_0 \in [0, 1)$ such that every admissible input and disturbance trajectory in the domain of Assumption 1 satisfies

$$\|\mathbf{I}_p - \mathbf{S}_t(\mathbf{U}, \mathbf{d})\mathbf{K}_0\|_2 \leq r_0 < 1, \quad t = 1, \dots, T. \quad (62)$$

Lemma 4.2 Under Assumptions 1 and 3, the family $\mathbf{A}_k := \mathbf{I} - \mathbf{G}_k \mathbf{L}$ admits a common weighted Euclidean norm in which it is uniformly contractive.

Proof Causality makes $\mathbf{G}_k$ block lower triangular; $\mathbf{L}$ is block lower bidiagonal with diagonal block $\mathbf{K}_0$. Thus $\mathbf{A}_k$ is block lower triangular. For its $p \times p$ diagonal blocks,

$$\mathbf{A}_{k,tt} = \mathbf{I}_p - \mathbf{G}_{k,tt}\mathbf{K}_0, \quad \mathbf{G}_{k,tt} = \int_0^1 \mathbf{S}_t(\mathbf{U}_k + s\Delta\mathbf{U}_k, \mathbf{d}_k) ds. \quad (63)$$

The input segment lies in the admissible box, so Assumption 3 and convexity of the spectral norm imply $\|\mathbf{A}_{k,tt}\|_2 \leq r_0$.

For $j = 1, \dots, T-1$, define the finite uniform bounds

$$M_j := \sup_{\mathbf{A} \in \mathcal{A}} \max_{i=j+1, \dots, T} \|\mathbf{A}_{i,i-j}\|_2, \quad \mathcal{A} := \{\mathbf{I} - \mathbf{G}(\mathbf{U}, \mathbf{U}', \mathbf{d})\mathbf{L} : \mathbf{U}, \mathbf{U}' \in \Omega_{\varepsilon_u}, \mathbf{d} \in \mathcal{D}\}, \quad (64)$$

where $\mathbf{G}(\mathbf{U}, \mathbf{U}', \mathbf{d})$ is the averaged Jacobian along the segment from $\mathbf{U}$ to $\mathbf{U}'$. These bounds are finite because the horizon is finite and the admissible matrix family is compact. Choose a fixed $\vartheta \in (0, 1)$ and a bound $\rho_\vartheta \in (0, 1)$ so that

$$r_0 + \sum_{j=1}^{T-1} \vartheta^j M_j \leq \rho_\vartheta < 1, \quad \mathbf{W}_\vartheta := \text{diag}(1, \vartheta, \dots, \vartheta^{T-1}) \otimes \mathbf{I}_p. \quad (65)$$

Such a choice exists because the finite sum tends to zero as $\vartheta \downarrow 0$. For $T = 1$, take $\mathbf{W}_\vartheta = \mathbf{I}_p$ and any $\rho_\vartheta \in (r_0, 1)$. A similarity transform by $\mathbf{W}_\vartheta$ multiplies the j th block subdiagonal by ϑ^j . The norm of one such block-shift operator is at most M_j ; the triangle inequality therefore gives, uniformly over the admissible family,

$$\|\mathbf{W}_\vartheta \mathbf{A}_k \mathbf{W}_\vartheta^{-1}\|_2 \leq r_0 + \sum_{j=1}^{T-1} \vartheta^j M_j \leq \rho_\vartheta < 1. \quad (66)$$

The same weight matrix is used for every iteration. $\square$

Theorem 4.2 Suppose Assumptions 1–3 hold and the initial learning input satisfies Theorem 4.1. Let $\|\mathbf{E}\|_{\vartheta} := \|\mathbf{W}_{\vartheta}\mathbf{E}\|_2$ and $\bar{\omega}_{\vartheta} := \sup_{k \geq k_0} \|\mathbf{W}_{\vartheta}\boldsymbol{\omega}_k\|_2$. Then, for $k \geq k_0$,

$$\|\mathbf{E}_k\|_{\vartheta} \leq \rho_{\vartheta}^{k-k_0} \|\mathbf{E}_{k_0}\|_{\vartheta} + \frac{1 - \rho_{\vartheta}^{k-k_0}}{1 - \rho_{\vartheta}} \bar{\omega}_{\vartheta}. \quad (67)$$

With $C_{\vartheta} := \|\mathbf{W}_{\vartheta}^{-1}\|_2 = \vartheta^{-(T-1)}$ (and $C_{\vartheta} = 1$ for $T = 1$),

$$\|\mathbf{E}_k\|_2 \leq C_{\vartheta} \rho_{\vartheta}^{k-k_0} \|\mathbf{E}_{k_0}\|_2 + C_{\vartheta} \frac{1 - \rho_{\vartheta}^{k-k_0}}{1 - \rho_{\vartheta}} \bar{\omega}_{\vartheta}, \quad (68)$$

and hence

$$\limsup_{k \rightarrow \infty} \|\mathbf{E}_k\|_2 \leq \frac{C_{\vartheta} \bar{\omega}_{\vartheta}}{1 - \rho_{\vartheta}} \leq \frac{C_{\vartheta} \bar{\omega}}{1 - \rho_{\vartheta}} =: \epsilon_E. \quad (69)$$

Proof Multiplication of (54) by $\mathbf{W}_{\vartheta}$ and Lemma 4.2 give $\|\mathbf{E}_{k+1}\|_{\vartheta} \leq \rho_{\vartheta} \|\mathbf{E}_k\|_{\vartheta} + \bar{\omega}_{\vartheta}$. Iteration yields (67); norm equivalence gives the ordinary Euclidean estimates. Since $\bar{\omega}_{\vartheta} \leq \bar{\omega}$, (68) implies

$$\|\mathbf{E}_k\|_2 \leq C_{\vartheta} \rho_{\vartheta}^{k-k_0} \|\mathbf{E}_{k_0}\|_2 + \epsilon_E.$$

The constants C_{ϑ} , ρ_{ϑ} , and ϵ_E are independent of k_0 and the particular admissible trajectory. For any bounded set of initial errors and any radius greater than ϵ_E , the first term becomes smaller than the difference between this radius and ϵ_E after a uniformly bounded number of iterations. The tracking error therefore enters and remains within the corresponding ball, establishing UUB. $\square$

5 Numerical Simulations

5.1 Simulation Model and Reference Trajectory

The simulations consider a modified two-degree-of-freedom nonlinear mass–spring–damper system based on Example 2 in [40]. The equations of motion are

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} + \mathbf{f}_{\text{nl}}(\mathbf{q}) = \mathbf{u} + \mathbf{d}, \quad (70)$$

Here, $\mathbf{q}$ and $\dot{\mathbf{q}}$ denote the displacement and velocity vectors, measured in m and m/s, respectively. The vectors $\mathbf{u}$, $\mathbf{d}$, and $\mathbf{f}_{\text{nl}}(\mathbf{q})$ represent the applied control forces, disturbance forces, and nonlinear restoring forces, respectively, all measured in N. The matrices $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the mass, damping, and stiffness matrices, with entries expressed in kg, N s/m, and N/m, respectively. Their values are

$$\mathbf{M} = \begin{bmatrix} 0.25 & 0.00 \\ 0.00 & 0.20 \end{bmatrix} \text{ kg}, \quad \mathbf{C} = \begin{bmatrix} 0.25 & -0.06 \\ -0.06 & 0.22 \end{bmatrix} \text{ N s/m}, \quad (71)$$

$$\mathbf{K} = \begin{bmatrix} 0.80 & -0.20 \\ -0.20 & 0.70 \end{bmatrix} \text{ N/m}, \quad (72)$$

and

$$\begin{aligned} \mathbf{f}_{\text{nl}}(\mathbf{q}) = & \begin{bmatrix} 0.08 \sin(q_1/q_0) + 0.04(q_1/q_0)^3 \\ 0.06 \sin(q_2/q_0) + 0.035(q_2/q_0)^3 \end{bmatrix} \text{ N} \\ & + 0.03 \begin{bmatrix} \sin((q_1 - q_2)/q_0) \\ \sin((q_2 - q_1)/q_0) \end{bmatrix} \text{ N}, \quad q_0 = 1 \text{ m}. \end{aligned} \quad (73)$$

The reference length scale $q_0 = 1$ m makes the arguments of the trigonometric functions dimensionless. The state is $\mathbf{x}_k(t) = [\mathbf{q}_k^\top(t), \dot{\mathbf{q}}_k^\top(t)]^\top \in \mathbb{R}^4$, where $\mathbf{q}_k(t), \dot{\mathbf{q}}_k(t) \in \mathbb{R}^2$ denote the displacement and velocity vectors, respectively. For this simulation, $n = 4$ and $m = p = n_d = 2$. The primary and auxiliary outputs are $\mathbf{y}_k(t) = \mathbf{q}_k(t)$ and $\mathbf{v}_k(t) = \dot{\mathbf{q}}_k(t)$, respectively. Thus, $\mathbf{y}_k, \mathbf{e}_k$ are measured in m, whereas $\mathbf{v}_k, \mathbf{e}_k^v$ are measured in m/s. The continuous system is discretized by fourth-order Runge–Kutta (RK4). Its one-step map f_{RK4} serves as f in (1) and is continuously differentiable in state and input on the considered operating region; the local Jacobians in Sect. 4 are those of this discrete transition.

Let $\tau = tT_s$ denote physical time, measured in s. The reference trajectories and disturbances below are expressed in τ . The desired displacement and velocity trajectories

and initial state are defined as

$$\begin{aligned}
\mathbf{y}_d(\tau) &= \begin{bmatrix} \cos(0.8\tau) \\ \sin(0.7\tau) \end{bmatrix} \text{ m}, \\
\mathbf{v}_d(\tau) = \dot{\mathbf{y}}_d(\tau) &= \begin{bmatrix} -0.8 \sin(0.8\tau) \\ 0.7 \cos(0.7\tau) \end{bmatrix} \text{ m/s}, \\
\mathbf{q}_k(0) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ m}, \quad \dot{\mathbf{q}}_k(0) = \begin{bmatrix} 0 \\ 0.7 \end{bmatrix} \text{ m/s}.
\end{aligned} \tag{74}$$

Both reference displacements have an amplitude of 1 m, with angular frequencies 0.8 rad/s and 0.7 rad/s, respectively.

The initial state in (74) matches the reference displacement and velocity at $\tau = 0$, giving $\mathbf{e}_k(0) = \mathbf{e}_k^v(0) = 0$. To compare the input required by the reference trajectory with the actuator limits, consider the disturbance-free nominal inverse-dynamics input $\mathbf{u}_d = \mathbf{M}\ddot{\mathbf{y}}_d + \mathbf{C}\dot{\mathbf{y}}_d + \mathbf{K}\mathbf{y}_d + \mathbf{f}_{\text{nl}}(\mathbf{y}_d)$. Over the horizon, the componentwise minima and maxima of $\mathbf{u}_d$ are $[-0.8971, -0.8797]^\top \text{ N}$ and $[0.9551, 0.8095]^\top \text{ N}$, respectively. Both maxima exceed the corresponding upper input limit of 0.8 N.

5.2 Simulation Settings

The disturbance in each iteration is

$$d_{k,i}(\tau) = A_{k,i} \sin(\omega_{d,i}\tau + \varphi_{k,i}) + c_{k,i} + (0.003 \text{ N})\nu_{k,i}(\tau), \tag{75}$$

The disturbance amplitudes $A_{k,i}$ and offsets $c_{k,i}$ are sampled uniformly from $[0.04, 0.08] \text{ N}$ and $[-0.006, 0.006] \text{ N}$, respectively. The phases $\varphi_{k,i}$ are sampled uniformly from $[-\pi, \pi] \text{ rad}$, and the angular frequencies are $[\omega_{d,1}, \omega_{d,2}]^\top =$

$[1.2, 0.9]^\top$ rad/s. The dimensionless standard Gaussian samples $\nu_{k,i}$ are truncated to $[-2.5, 2.5]$. Amplitudes, phases, and offsets are resampled each iteration.

For the simulations, $\mathbf{s}_k(t) \in \mathbb{R}^{66}$ comprises five blocks: core (16), history (32), time (3), input margins (2), and cross-iteration information (13). Continuous features are normalized using fixed scales or actuator bounds. Time features and the validity indicator are used without additional scaling.

Table 2 lists the main simulation and controller parameters. PD-ILC + HC, PD-ILC + Two-Layer Barrier, and the proposed method use the same $\mathbf{K}_p$ and $\mathbf{K}_v$ throughout the comparison. FB-ILC [32] is included as a comparison method from the literature. All methods use the same plant, reference, disturbance realizations, input bounds, and horizon. The overbarred quantities in (31) are normalized channelwise using the reward scales in Table 2; for example, $\bar{\mathbf{e}}_k = \mathbf{e}_k \oslash \mathbf{s}_e$ and $\bar{\mathbf{u}}_k^S = \mathbf{u}_k^S \oslash \mathbf{s}_u$, where $\oslash$ denotes elementwise division.

The SAC policy and critic networks each contain two hidden layers with 256 ReLU units. Both the actor and critics use a learning rate of 10^{-4} , with discount factor $\gamma = 0.995$ and target-network update rate $\tau_{\text{soft}} = 0.01$.

SAC training comprises two successive stages of 500 iterations each. The ILC learning sequence and stored data from the preceding iteration are reset at the start of each stage, while the networks and replay buffer are retained. Both policy candidates are validated on the same 200 held-out disturbance realizations. The second-stage policy is selected for its lower validation root-mean-square error (RMSE) ratio relative to the corresponding controller with the residual action set to zero. The methods are evaluated over $K_{\text{eval}} = 500$ iterations with frozen network parameters and a deterministic policy. The reported RMSE statistics are the mean and standard deviation over the final 50 evaluation iterations.

Table 2 Main simulation and controller parameters

Parameter	Value	Unit
T_s	0.01	s
T	800	—
$\mathbf{u}_{\min}$	$[-0.9, -0.9]^\top$	N
$\mathbf{u}_{\max}$	$[0.8, 0.8]^\top$	N
$\mathbf{K}_p$	$\begin{bmatrix} 0.052 & -0.0106 \\ -0.0106 & 0.045 \end{bmatrix}$	N/m
$\mathbf{K}_v$	$\begin{bmatrix} 0.085 & -0.0198 \\ -0.0198 & 0.078 \end{bmatrix}$	N s/m
ε_u	5×10^{-5}	N
κ_1	5×10^{-5}	N ²
κ_2	5×10^{-5}	N ²
$\bar{\mathbf{a}}$	$[0.08, 0.08]^\top$	N
q_e	12	—
q_v	0.10	—
q_u	0.005	—
q_s	0.02	—
q_b	0.05	—
q_c	0.40	—
$\mathbf{s}_e$	$[0.10, 0.10]^\top$	m
$\mathbf{s}_v$	$[0.20, 0.20]^\top$	m/s
$\mathbf{s}_u$	$[0.08, 0.08]^\top$	N
$\mathbf{s}_\zeta$	$[0.08, 0.08]^\top$	N
$\mathbf{s}_{\delta u}$	$[0.04, 0.04]^\top$	N

The performance metrics are defined as

$$R_k = \sqrt{\frac{1}{pT} \sum_{t=1}^T \|\mathbf{e}_k(t)\|_2^2}, \quad (76)$$

$$\mu_k = \frac{\|\mathbf{U}_k^I\|_2}{\sqrt{mT}}, \quad (77)$$

$$\chi_k = \|\mathbf{U}_k - \mathbf{C}_k\|_\infty. \quad (78)$$

Here, R_k is measured in m, while μ_k and χ_k are measured in N. For each compared scheme, χ_k quantifies the maximum difference between the corresponding pre-constraint command and the actually applied control input.

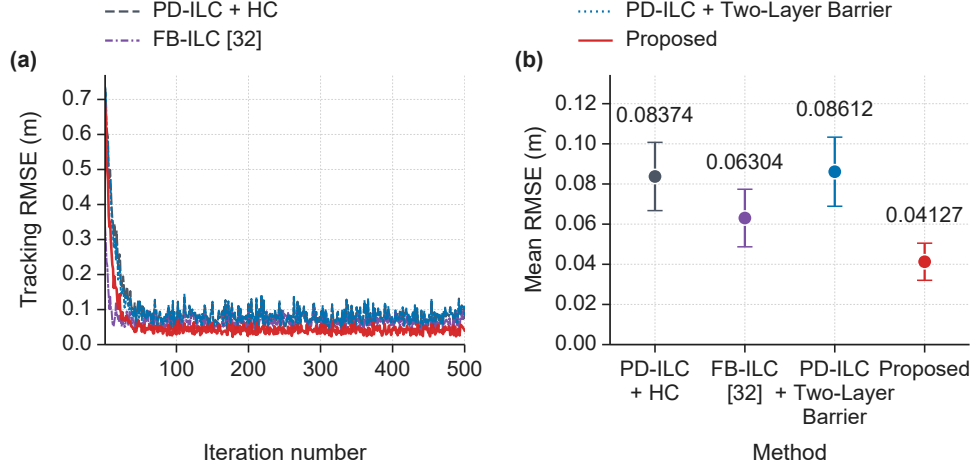


Fig. 2 Iteration-wise tracking performance of the compared methods: (a) RMSE evolution; (b) mean RMSE over the final 50 evaluation iterations, with error bars showing one standard deviation

5.3 Tracking Performance

Four methods are compared: PD-ILC with hard clipping (PD-ILC + HC), FB-ILC [32], PD-ILC + Two-Layer Barrier, and the proposed method. PD-ILC + HC applies hard clipping to the plant command while retaining the stored PD-ILC learning update. PD-ILC + Two-Layer Barrier uses the two barrier mappings with the SAC residual action set to zero, whereas the proposed method incorporates the validated SAC residual compensator.

Over the final 50 evaluation iterations, the mean RMSE values are 0.08374 m, 0.06304 m, 0.08612 m, and 0.04127 m for PD-ILC + HC, FB-ILC [32], PD-ILC + Two-Layer Barrier, and the proposed method, respectively (Fig. 2). The proposed method reduces mean RMSE by 50.72%, 34.54%, and 52.08% relative to the three comparison methods, respectively.

The final-iteration tracking responses are reported in Fig. 3.

5.4 Input-Constraint Handling

In the displayed final iteration, all four schemes have a 0% input-constraint violation rate in both control-input channels (Fig. 4).

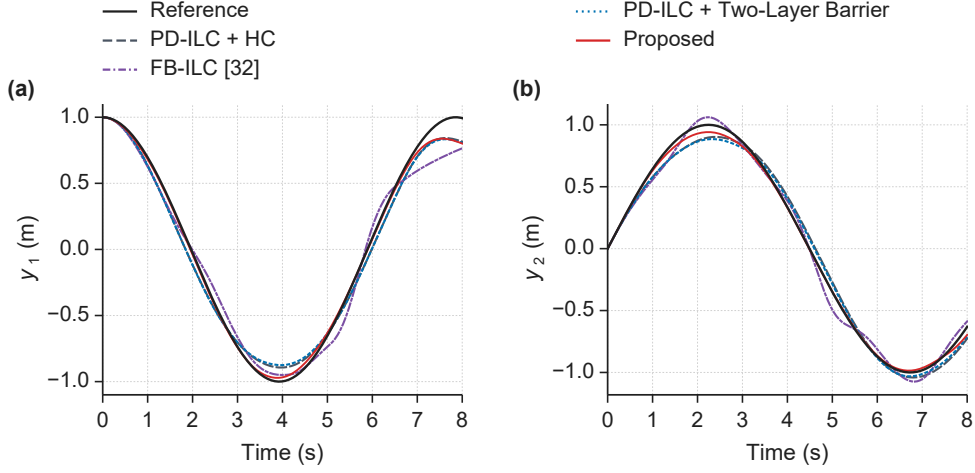


Fig. 3 Final-iteration trajectory tracking responses of the compared methods

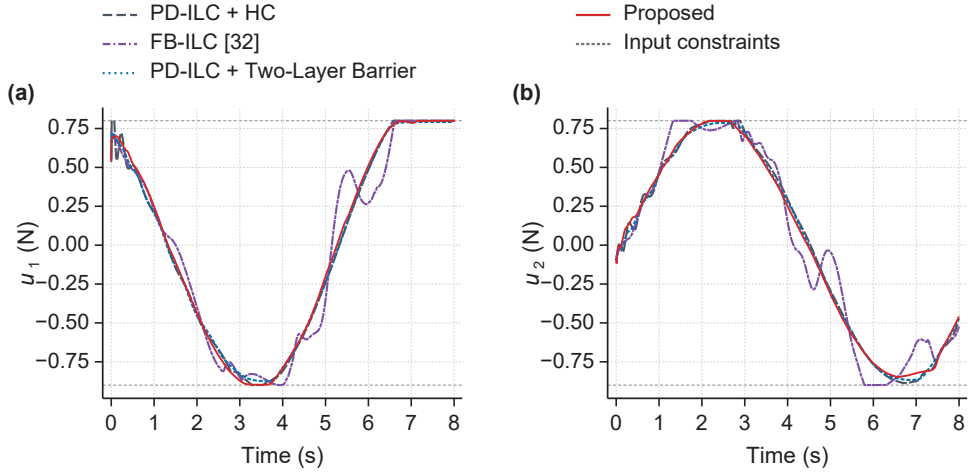


Fig. 4 Applied control inputs and input-constraint boundaries in the final iteration

For the three comparison configurations without SAC residual compensation, the mean values of μ_k over the final 50 evaluation iterations are 2.126, 0.630, and 0.611 for PD-ILC + HC, FB-ILC [32], and PD-ILC + Two-Layer Barrier, respectively (Fig. 5(a)). Hard clipping constrains the applied control input while the stored learning input can accumulate outside the feasible set. The first-layer mapping constrains this stored sequence. The mean values of χ_k over the final 50 evaluation iterations are

8.864, 0.209 N, 0.006, and 0.073 for PD-ILC + HC, FB-ILC [32], PD-ILC + Two-Layer Barrier, and the proposed method, respectively (Fig. 5(b)).

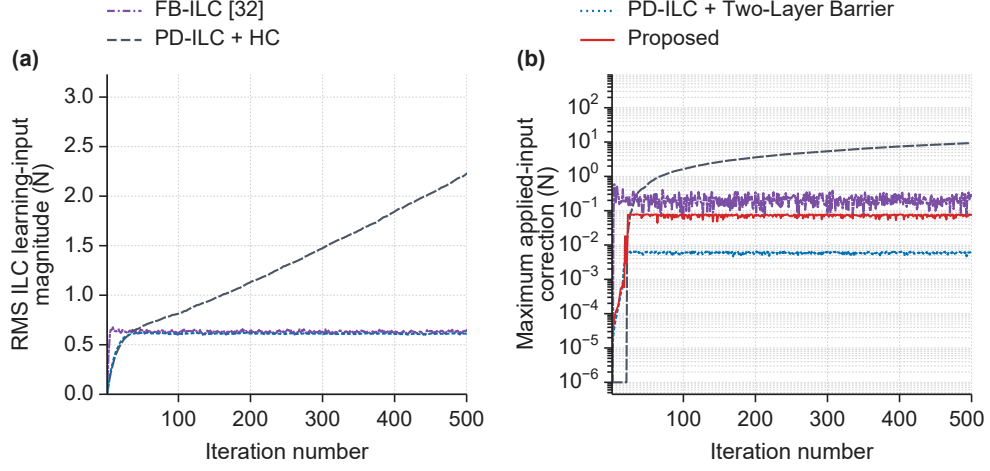


Fig. 5 Evolution of the learning-input magnitude and applied-input correction: (a) root-mean-square (RMS) ILC learning-input magnitude μ_k for the three schemes without SAC; (b) maximum applied-input correction χ_k for all four schemes

5.5 Effects of the Two Barrier Mappings

With the SAC residual action fixed at zero, the average RMSE across the four barrier configurations remains between 0.080 and 0.086 (Fig. 6(a)). The stored learning-input violation rate in Fig. 6(b) is computed over the final $K_{\text{tail}} = 50$ evaluation iterations, $T = 800$ time samples, and both input channels. Define $\mathcal{K}_{\text{tail}} = \{K_{\text{eval}} - K_{\text{tail}}, \dots, K_{\text{eval}} - 1\}$ and

$$\begin{aligned} \iota_i(u) &:= \mathbf{1}\{u < u_{\min,i} - \delta_{\text{num}} \text{ or } u > u_{\max,i} + \delta_{\text{num}}\}, \\ P_I &:= \frac{100}{mTK_{\text{tail}}} \sum_{k \in \mathcal{K}_{\text{tail}}} \sum_{t=0}^{T-1} \sum_{i=1}^m \iota_i(u_{k+1,i}^I(t)), \end{aligned} \quad (79)$$

where $\delta_{\text{num}} = 10^{-10}$ matches the numerical violation tolerance. The sequence $\mathbf{U}_{k+1}^I$ is stored after the update of iteration k ; the test uses the physical bounds $\mathbf{u}_{\min}, \mathbf{u}_{\max}$

rather than the tightened barrier bounds. Without either barrier layer and with layer 2 alone, P_I is 13.1% and 16.9%, respectively. Enabling layer 1 reduces the stored learning-input violation rate to 0%.

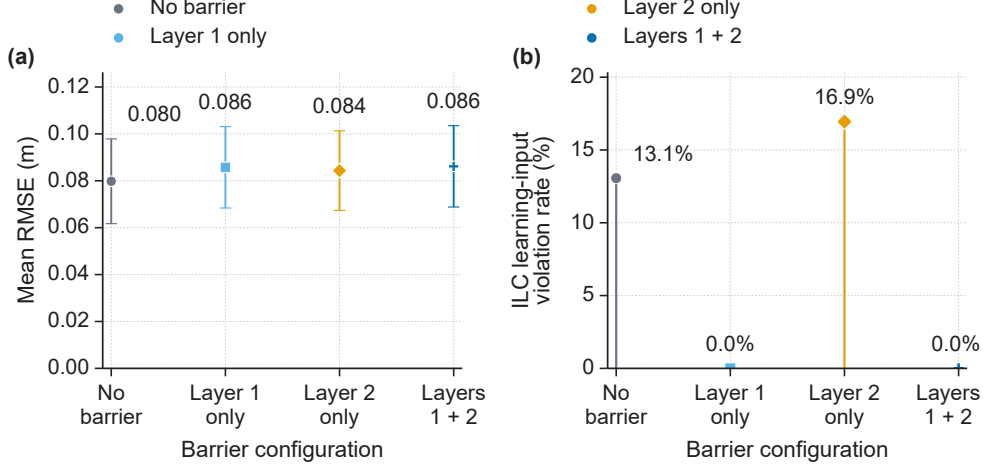


Fig. 6 Ablation study of learning-input constraints: (a) mean tracking RMSE over the final 50 evaluation iterations, with error bars showing one standard deviation; (b) stored learning-input violation rate over the final 50 evaluation iterations, aggregated across all time samples and both input channels

Figure 7 compares the second-layer input correction and examines input feasibility with a prescribed bounded compensation input. For Fig. 7(a), $J_b = K_{\text{eval}}^{-1} \sum_{k=0}^{K_{\text{eval}}-1} \|\mathbf{U}_k - \mathbf{C}_k\|_2$ is the mean finite-horizon norm of the second-layer correction across all evaluation iterations. With layer 2 alone, $J_b = 45.7 \text{ N}$; with both layers active, it is 0.0701 N , a reduction of approximately 99.8%. For this test, a prescribed deterministic compensation input is added to the final stored learning input obtained with layer 1 only. The compensation input is directed toward the nearer physical input bound. At samples whose distance to the nearest physical bound is at most $1.25\bar{a}_i$, the compensation magnitude is $\bar{a}_i = 0.08 \text{ N}$ per channel; it is zero elsewhere. Over $T = 800$ samples in the displayed final iteration, the pre-barrier signal violates the bounds at rates of 32.0% and 28.6% in the two channels with layer 1 alone. Applying layer 2 reduces both rates to 0%.

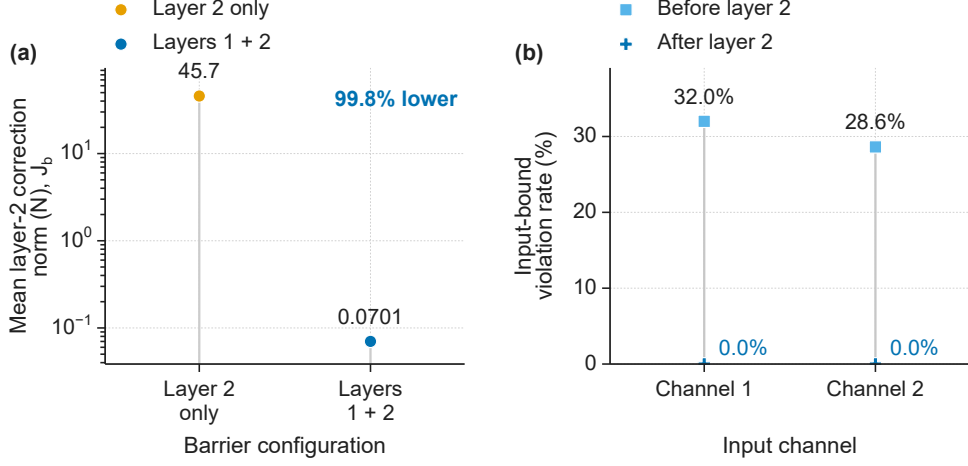


Fig. 7 Ablation study of applied-input constraints: (a) mean finite-horizon norm J_b of the second-layer correction over all evaluation iterations; (b) channelwise input-bound violation rates before and after the second-layer mapping with the prescribed compensation input added to the final stored learning input

5.6 SAC Training and Residual Compensation

Figure 8 shows the mean per-step reward over two successive SAC training stages of 500 iterations each. The reward increased rapidly from its initial level in each training stage and subsequently fluctuated around a higher plateau. The drop at the start of Stage 2 coincided with the reset of the ILC learning sequence and stored data from the preceding iteration. The mean rewards over the final 50 training iterations were -7.274 and -5.614 for Stages 1 and 2, respectively.

The effective applied residual input after the second-layer barrier mapping is defined relative to the stored ILC learning input as

$$\mathbf{u}_k^{\text{res}}(t) = \mathbf{u}_k(t) - \mathbf{u}_k^I(t) = \mathbf{u}_k^S(t) + \boldsymbol{\zeta}_k(t). \quad (80)$$

The channelwise correlation and RMS mismatch in (81) characterize the relation between the effective applied residual input $\mathbf{u}_k^{\text{res}}$ and the disturbance:

$$\varrho_i = \text{corr}(u_{k,i}^{\text{res}}, -d_{k,i}), \quad R_i^c = \text{RMS}(u_{k,i}^{\text{res}} + d_{k,i}). \quad (81)$$

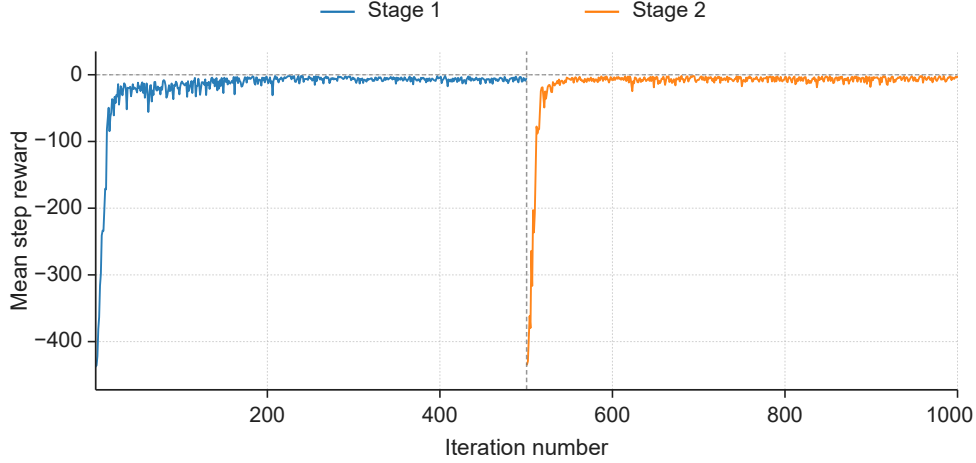


Fig. 8 Mean per-step SAC reward over two successive 500-iteration training stages. The dashed line marks the transition between stages

For each evaluation iteration, these statistics are computed using all $T = 800$ time samples in each channel. Over the final 50 evaluation iterations, the channelwise correlations between the effective applied residual input and the negative disturbance average 0.579 and 0.662. The corresponding RMS mismatches are 0.0408 N and 0.0464 N, while the mean channelwise SAC-action RMS values are 0.0459 N and 0.0510 N. The positive correlations indicate that the effective applied residual input tends to oppose disturbance variations in both channels.

6 Conclusion

This paper developed a SAC-enhanced PD-ILC scheme with two recentered barrier mappings for nonlinear MIMO repetitive systems under nonrepetitive disturbances and input constraints. The mappings maintain interior feasibility of the stored learning sequence and the applied input. An exact error recursion was derived from the nonlinear finite-horizon input-output map. Under the stated assumptions, an explicit tracking-error bound was derived in a common weighted norm, establishing uniform ultimate boundedness.

Simulations on the coupled nonlinear mechanical system yielded a mean RMSE of 0.04127 m over the final 50 evaluation iterations, a 52.08% reduction relative to PD-ILC + Two-Layer Barrier. Both applied-input channels had a 0% constraint-violation rate in the displayed final iteration. The ablations showed that layer 1 prevented stored learning-input violations, while layer 2 restored applied-input feasibility under the prescribed bounded compensation input.

References

- [1] Bristow, D.A., Tharayil, M., Alleyne, A.G.: A survey of iterative learning control. *IEEE Control Syst. Mag.* **26**(3), 96–114 (2006)
<https://doi.org/10.1109/MCS.2006.1636313>
- [2] Stadler, G., Steinboeck, A., Marko, L., Deutschmann-Olek, A., Kugi, A.: Iterative learning and feedback control for the curvature and contact force of a metal strip on a roll. *Control Eng. Pract.* **121**, 105071 (2022)
<https://doi.org/10.1016/j.conengprac.2022.105071>
- [3] Yang, L., Li, Y., Huang, D., Xia, J., Zhou, X.: Spatial iterative learning control for robotic path learning. *IEEE Trans. Cybern.* **52**(7), 5789–5798 (2022)
<https://doi.org/10.1109/TCYB.2021.3138992>
- [4] Wu, W., Qiu, L., Liu, X., Guo, F., Rodriguez, J., Ma, J., Fang, Y.: Data-driven iterative learning predictive control for power converters. *IEEE Trans. Power Electron.* **37**(12), 14028–14033 (2022)
<https://doi.org/10.1109/TPEL.2022.3194518>
- [5] Boudjedir, C.E., Boukhetala, D., Bouri, M.: Iterative learning control of multivariable uncertain nonlinear systems with nonrepetitive trajectory. *Nonlinear Dyn.* **95**, 2197–2208 (2019)

<https://doi.org/10.1007/s11071-018-4685-0>

- [6] Boudjedir, C.E., Bouri, M., Boukhetala, D.: Model-free iterative learning control with nonrepetitive trajectories for second-order MIMO nonlinear systems—application to a Delta robot. *IEEE Trans. Ind. Electron.* **68**(8), 7433–7443 (2021) <https://doi.org/10.1109/TIE.2020.3007091>
- [7] Chi, R., Hui, Y., Chien, C.-J., Huang, B., Hou, Z.: Convergence analysis of sampled-data ILC for locally Lipschitz continuous nonlinear nonaffine systems with nonrepetitive uncertainties. *IEEE Trans. Autom. Control* **66**(7), 3347–3354 (2021) <https://doi.org/10.1109/TAC.2020.3020803>
- [8] Li, S., Li, X.: Finite-time extended state observer-based iterative learning control for nonrepeatable nonlinear systems. *Nonlinear Dyn.* **113**, 16531–16543 (2025) <https://doi.org/10.1007/s11071-025-11016-3>
- [9] Xing, J., Chi, R., Lin, N.: Adaptive iterative learning control for 2D nonlinear systems with nonrepetitive uncertainties. *Int. J. Robust Nonlinear Control* **31**(4), 1168–1180 (2021) <https://doi.org/10.1002/rnc.5347>
- [10] Meng, D., Zhang, J.: Robust optimization-based iterative learning control for nonlinear systems with nonrepetitive uncertainties. *IEEE/CAA J. Autom. Sinica* **8**(5), 1001–1014 (2021) <https://doi.org/10.1109/JAS.2021.1003973>
- [11] He, C., Li, J., Liu, S., Wang, J.: Robust model-based predictive iterative learning control for systems with non-repetitive disturbances. *Nonlinear Anal. Hybrid Syst.* **51**, 101436 (2024) <https://doi.org/10.1016/j.nahs.2023.101436>
- [12] Smith, R.D., Alleyne, A.G.: Set-to-set iterative learning control for non-repetitive disturbances. *IEEE Control Syst. Lett.* **9**, 2441–2446 (2025) <https://doi.org/10.1109/LCSYS.2025.3624185>

- [13] Hao, S., Liu, T., Paszke, W.: PIO based data-driven iterative learning control for nonlinear batch processes with nonrepetitive disturbances subject to input constraints. *IFAC-PapersOnLine* **54**(3), 25–30 (2021)
<https://doi.org/10.1016/j.ifacol.2021.08.213>
- [14] Tao, H., Zheng, J., Wei, J., Paszke, W., Rogers, E., Stojanovic, V.: Repetitive process based indirect-type iterative learning control for batch processes with model uncertainty and input delay. *J. Process Control* **132**, 103112 (2023)
<https://doi.org/10.1016/j.jprocont.2023.103112>
- [15] Lin, N., Chi, R., Huang, B., Hou, Z.: Iterative dynamic linearization and identification of a nonlinear learning controller: A data-driven approach. *J. Franklin Inst.* **356**(13), 7009–7027 (2019)
<https://doi.org/10.1016/j.jfranklin.2019.06.011>
- [16] Yu, X., Hou, Z., Polycarpou, M.M., Duan, L.: Data-driven iterative learning control for nonlinear discrete-time MIMO systems. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(3), 1136–1148 (2021)
<https://doi.org/10.1109/TNNLS.2020.2980588>
- [17] Chen, Y., Jiang, W., Charalambous, T.: Machine learning based iterative learning control for non-repetitive time-varying systems. *Int. J. Robust Nonlinear Control* **33**(7), 4098–4116 (2023) <https://doi.org/10.1002/rnc.6272>
- [18] Lee, Y.-H., Rai, S., Tsao, T.-C.: Data-driven iterative learning control of nonlinear systems by adaptive model matching. *IEEE/ASME Trans. Mechatronics* **27**(6), 5626–5636 (2022)
<https://doi.org/10.1109/TMECH.2022.3176984>
- [19] Liu, J., Zhou, Z., Hong, W., Shi, J.: Two-dimensional iterative learning control

- with deep reinforcement learning compensation for the non-repetitive uncertain batch processes. *J. Process Control* **131**, 103106 (2023)
<https://doi.org/10.1016/j.jprocont.2023.103106>
- [20] Soleimani, E., Sedigh, A.K., Nikoofard, A.: Data-driven reinforcement learning-based forgetting factor iterative learning control. *IEEE Trans. Autom. Sci. Eng.* **22**, 12245–12256 (2025) <https://doi.org/10.1109/TASE.2025.3540699>
- [21] Tao, H., Huang, Y., Liu, T., Paszke, W.: Reinforcement learning based iterative learning control for nonlinear batch process with non-repetitive uncertainty via Koopman operator. *J. Process Control* **148**, 103402 (2025)
<https://doi.org/10.1016/j.jprocont.2025.103402>
- [22] Zheng, Y., Tao, J., Sun, Q., Sun, H., Chen, Z., Sun, M., Xie, G.: Soft actor–critic based active disturbance rejection path following control for unmanned surface vessel under wind and wave disturbances. *Ocean Eng.* **247**, 110631 (2022) <https://doi.org/10.1016/j.oceaneng.2022.110631>
- [23] Chu, B., Owens, D.H.: Iterative learning control for constrained linear systems. *Int. J. Control* **83**(7), 1397–1413 (2010)
<https://doi.org/10.1080/00207171003758752>
- [24] Liao-McPherson, D., Balta, E.C., Rupenyan, A., Lygeros, J.: On robustness in optimization-based constrained iterative learning control. *IEEE Control Syst. Lett.* **6**, 2846–2851 (2022) <https://doi.org/10.1109/LCSYS.2022.3178877>
- [25] Chi, R., Liu, X., Zhang, R., Hou, Z., Huang, B.: Constrained data-driven optimal iterative learning control. *J. Process Control* **55**, 10–29 (2017)
<https://doi.org/10.1016/j.jprocont.2017.03.003>
- [26] Yu, Q., Hou, Z., Bu, X., Yang, J.: Observer-based data-driven constrained norm

- optimal iterative learning control for unknown non-affine non-linear systems with both available and unavailable system states. *J. Franklin Inst.* **357**(10), 5852–5877 (2020) <https://doi.org/10.1016/j.jfranklin.2020.03.019>
- [27] Chu, B., Rauh, A., Aschemann, H., Rogers, E., Owens, D.H.: Constrained iterative learning control for linear time-varying systems with experimental validation on a high-speed rack feeder. *IEEE Trans. Control Syst. Technol.* **30**(5), 1834–1846 (2022) <https://doi.org/10.1109/TCST.2021.3123744>
- [28] Yu, Q., Hou, Z., Chi, R.: Adaptive iterative learning control for nonlinear uncertain systems with both state and input constraints. *J. Franklin Inst.* **353**(15), 3920–3943 (2016) <https://doi.org/10.1016/j.jfranklin.2016.07.007>
- [29] He, W., Meng, T., He, X., Ge, S.S.: Unified iterative learning control for flexible structures with input constraints. *Automatica* **96**, 326–336 (2018) <https://doi.org/10.1016/j.automatica.2018.06.051>
- [30] Zhang, J., Meng, D.: Convergence analysis of saturated iterative learning control systems with locally Lipschitz nonlinearities. *IEEE Trans. Neural Netw. Learn. Syst.* **31**(10), 4025–4035 (2020) <https://doi.org/10.1109/TNNLS.2019.2951752>
- [31] Xu, J.-X., Tan, Y., Lee, T.H.: Iterative learning control design based on composite energy function with input saturation. *Automatica* **40**(8), 1371–1377 (2004) <https://doi.org/10.1016/j.automatica.2004.01.029>
- [32] Sebastian, G., Tan, Y., Oetomo, D.: Convergence analysis of feedback-based iterative learning control with input saturation. *Automatica* **101**, 44–52 (2019) <https://doi.org/10.1016/j.automatica.2018.11.045>
- [33] Song, Z., Gao, L., Wang, Z., Li, P.: Adaptive neural control of constrained MIMO nonlinear systems with asymmetric input saturation and dead zone.

- IEEE Trans. Neural Netw. Learn. Syst. **35**(12), 18771–18783 (2024)
<https://doi.org/10.1109/TNNLS.2023.3321596>
- [34] Zhuang, Z., Tao, H., Chen, Y., Rogers, E., Oomen, T., Paszke, W.: Alternating projection-based iterative learning control for discrete-time systems with non-uniform trial lengths. *Int. J. Robust Nonlinear Control* **33**(12), 7333–7356 (2023) <https://doi.org/10.1002/rnc.6750>
- [35] Ames, A.D., Xu, X., Grizzle, J.W., Tabuada, P.: Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Trans. Autom. Control* **62**(8), 3861–3876 (2017) <https://doi.org/10.1109/TAC.2016.2638961>
- [36] Tee, K.P., Ge, S.S., Tay, E.H.: Barrier Lyapunov functions for the control of output-constrained nonlinear systems. *Automatica* **45**(4), 918–927 (2009) <https://doi.org/10.1016/j.automatica.2008.11.017>
- [37] Wills, A.G., Heath, W.P.: Barrier function based model predictive control. *Automatica* **40**(8), 1415–1422 (2004) <https://doi.org/10.1016/j.automatica.2004.03.002>
- [38] Haarnoja, T., Zhou, A., Abbeel, P., Levine, S.: Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: *Proc. 35th Int. Conf. Mach. Learn.*, vol. 80, pp. 1861–1870 (2018)
- [39] Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., Levine, S.: Soft Actor-Critic Algorithms and Applications. *arXiv preprint arXiv:1812.05905* (2018) <https://doi.org/10.48550/arXiv.1812.05905>
- [40] Liu, Y.-J., Tong, S.-C., Li, T.-S.: Observer-based adaptive fuzzy tracking control for a class of uncertain nonlinear MIMO systems. *Fuzzy Sets and*

Statements and Declarations

Funding. This work was partially supported by the National Natural Science Foundation of China under Grants 62361136585 and 62603297, partially by the 111 Project under Grant B23008, partially by the National Science Centre Poland under Grant 2023/48/Q/ST7/00205, and partially by the Fundamental Research Funds for the Central Universities under Grant JUSRP202601030.

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Author contributions. **Hongfeng Tao:** Supervision, including guidance on research planning and execution and oversight of manuscript drafting.

Fei Shao: Writing – original draft, Software, Visualization.

Zhihe Zhuang: Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Conceptualization.

Wojciech Paszke: Methodology, Supervision, Resources.

Data availability. The datasets generated during the current study are available from the corresponding author on reasonable request.

Graphical abstract

