Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(7): 23308–23343.
DOI: 10.3934/math.2026940
Received: 12 May 2026
Revised: 21 July 2026
Accepted: 22 July 2026
Published: 31 July 2026

Research article

A physics-informed Chebyshev spectral neural network for distributed-order fractional optimal control problems

Ishtiaq Ali *

Department of Mathematics and Statistics, College of Science, King Faisal University, P. O. Box 400, Al-Ahsa 31982, Saudi Arabia

* **Correspondence:** Email: iamirzada@kfu.edu.sa.

Abstract: In this paper, a physics-informed Chebyshev spectral neural network (PI-CSNN) framework was proposed for solving distributed-order fractional optimal control problems (DOFOCPs). The proposed approach combines Chebyshev spectral approximations with physics-informed residual minimization to construct an accurate computational framework for DOFOCPs. The distributed-order fractional operators are incorporated directly into the neural-spectral formulation, allowing the method to treat both single-order and distributed-order optimal control models within a unified framework. Several numerical examples involving solutions with limited regularity and time-invariant and time-varying DOFOCPs were presented to evaluate the performance of the proposed method. The numerical results demonstrated high accuracy, systematic convergence, and reliable approximations of the state, adjoint, and control variables. Comparisons with existing numerical methods showed that the proposed PI-CSNN framework achieves competitive optimal cost values and close agreement with benchmark solutions. Overall, the proposed PI-CSNN framework provides an accurate and flexible computational methodology for DOFOCPs.

Keywords: distributed-order fractional optimal control problems; physics-informed neural networks; Chebyshev spectral method; spectral neural approximation; distributed-order fractional differential equations
Mathematics Subject Classification: 26A33, 49M25, 65L60

1. Introduction

Distributed-order fractional differential equations (DOFDEs) have attracted increasing attention because they provide a flexible mathematical framework for describing dynamical systems with multiple memory scales and hereditary effects. Unlike classical integer-order models or single-order fractional formulations, DOFDEs involve a continuum of fractional orders and therefore provide a natural framework for modeling multiscale and history-dependent phenomena [1–3]. This capability has proved particularly valuable in viscoelasticity, anomalous diffusion, transport processes [4, 5], and ultraslow diffusion and multifractal dynamics, where a single fractional order is often insufficient to capture the underlying physics [6, 7]. Nevertheless, the continuous integration over the fractional-order variable introduces additional nonlocality and significantly increases the complexity of numerical approximation. Recent fast implicit difference schemes have improved the efficiency of solving time distributed-order and Riesz space-fractional diffusion-wave equations [8]; however, the accurate evaluation of distributed-order operators remains a major computational challenge, particularly when these operators are incorporated into coupled optimization problems.

These challenges become more pronounced for DOFOCPs, where the state, adjoint, and optimality equations must be solved simultaneously. Stable and convergent numerical schemes based on compact, finite difference, and reaction-diffusion formulations have been developed for distributed-order problems [9–11], while implicit finite difference, finite element, and multidimensional extensions have further broadened their applicability [12–14]. However, these methods have been developed primarily for forward distributed-order evolution equations rather than the coupled optimality systems arising in DOFOCPs. Such systems require the simultaneous treatment of left-sided state operators, right-sided adjoint operators, distributed-order quadrature, endpoint conditions, and control optimality relations. Consequently, the development of numerical methods that combine high-order accuracy with computational efficiency for distributed-order fractional optimal control problems remains an open and challenging research problem.

Operational matrix, collocation, wavelet, and spectral techniques provide an effective alternative by exploiting global basis representations to achieve high-order accuracy. For distributed-order fractional optimal control problems, Legendre collocation methods have demonstrated promising approximation properties [15], while Petrov-Galerkin and spectral collocation formulations have been successfully applied to distributed-order and multiterm fractional differential equations [16–18]. Chebyshev collocation, wavelet-based operational methods, and discontinuous Galerkin techniques have also been employed effectively for distributed-order models [19–21]. These studies demonstrate the effectiveness of high-order discretization techniques; however, they are generally based on directly solving the resulting algebraic systems for the unknown expansion coefficients and are primarily designed for forward distributed-order problems. Their extension to distributed-order optimal control problems remains computationally demanding because of the coupled state–adjoint system and the additional treatment of distributed-order operators. Moreover, relatively little attention has been given to the influence of spectral resolution, distributed-order quadrature, and solution regularity on the overall approximation accuracy. Physics-informed learning provides a complementary computational paradigm by enforcing the governing equations directly through the training objective, thereby eliminating the need for labeled solution data. However, conventional physics-informed neural networks approximate the solution variables directly in the physical domain and therefore do not explicitly exploit the high-order approximation properties of spectral basis functions. This motivates the development of a hybrid neural-spectral framework that combines the approximation efficiency of global Chebyshev expansions with the flexibility of physics-informed optimization.

Motivated by these considerations, this work develops a PI-CSNN framework for distributed-order fractional optimal control problems. The proposed framework combines boundary-adapted Chebyshev spectral approximations, physics-informed residual minimization, and Gauss-Legendre quadrature for distributed-order operators within a unified computational formulation. Unlike conventional physics-informed neural networks that approximate the solution variables directly in the physical domain, the proposed approach employs a neural feature extractor together with spectral output heads to generate the global Chebyshev coefficient vectors defining the state, adjoint, and, when required, control approximations. The prescribed initial and terminal conditions are incorporated exactly into the trial representations, while the coupled distributed-order optimality system is enforced through residual-based training without requiring labeled solution data. Furthermore, the proposed framework allows a systematic study of the effect of the spectral resolution, distributed-order quadrature, and regularity of the solution on the accuracy of the approximation and the convergence.

Classical neural network theory provides a powerful approximation framework for nonlinear functions and dynamical systems, forming the basis for modern neural network solvers [22, 23]. More recently, deep learning and physics-informed neural network approaches have demonstrated remarkable effectiveness for solving complex differential equations and related computational problems without relying on conventional mesh-based discretizations [24, 25]. Physics-informed learning reduces reliance on labeled solution data by embedding the governing equations into the training objective, maintaining consistency with the underlying mathematical model [26]. However, in most of the existing physics-informed neural networks, the unknown solution variables are directly approximated in the physical domain. For fractional-and distributed-order systems, this can lead to an increased computational cost due to the repeated evaluation of nonlocal operators. The training process can be further complicated by weak endpoint regularity and long-memory effects. Also, direct neural representations do not naturally inherit the high-order global approximation properties of orthogonal polynomial bases.

To overcome these limitations, several hybrid learning strategies have been proposed by combining neural networks with structured polynomial approximations. Polynomial-enhanced neural frameworks have been developed for differential and differential-algebraic systems [27], while Chebyshev neural approximations have demonstrated high accuracy for nonlinear differential equations [28]. In parallel, Bernoulli-barycentric rational collocation and implicit difference methods have improved the efficiency and accuracy of fractional differential equation solvers [29, 30]. Parallel-in-time algorithms and fast compact difference schemes have further accelerated large-scale fractional computations [31,32], while radial basis function neural networks have recently provided an efficient mesh-free alternative for diffusion equations [33]. Despite these developments, high-order spectral approximation, distributed-order operator evaluation, and physics-informed learning are typically treated as separate components rather than being integrated into a unified computational framework for distributed-order fractional optimal control problems. Physics-informed learning provides a natural mechanism for integrating governing equations directly into the training process. Existing physics-informed frameworks have demonstrated promising performance for fractional differential equations [34], while spectral collocation methods have achieved accurate approximations for nonlinear Riesz distributed-order fractional models [35]. Physics-informed neural methods have also been applied to fractional flow problems [36]. Nevertheless, existing approaches do not fully address the computational structure of distributed-order fractional optimal control problems, where the state equation, adjoint equation, control optimality condition, and distributed-order operators must be treated simultaneously within a unified computational framework.

Motivated by this observation, we develop a PI-CSNN framework for distributed-order fractional optimal control problems. The proposed framework combines boundary-adapted Chebyshev spectral approximations, Gauss-Legendre quadrature for distributed-order operators, and physics-informed residual minimization within a unified neural-spectral formulation. Unlike conventional physics-informed neural networks, which directly approximate the solution variables, the proposed approach employs a neural feature extractor and spectral output heads to generate the global Chebyshev coefficient vectors defining the state, adjoint, and, when required, control approximations. The coupled distributed-order optimality system is then enforced through residual-based optimization without requiring labeled solution data.

The main contribution of this work is the development of the PI-CSNN framework for distributed-order fractional optimal control problems. The proposed framework integrates boundary-adapted Chebyshev spectral approximations, Gauss-Legendre quadrature for the consistent evaluation of distributed-order Caputo and Riemann-Liouville operators, and physics-informed neural optimization within a unified computational framework. The boundary-adapted trial representations satisfy the prescribed state initial and adjoint terminal conditions exactly, while the governing state, adjoint, and optimality equations are enforced through residual-based training without requiring labeled solution data. The effectiveness of the proposed approach is demonstrated through numerical experiments on benchmark, time-invariant, and time-varying distributed-order fractional optimal control problems, illustrating its accuracy, convergence behavior, and robustness under different spectral resolutions and solution regularities.

The remainder of this paper is organized as follows. Section 2 reviews the preliminaries of distributed-order fractional calculus and Chebyshev spectral approximation. Section 3 presents the mathematical formulation of the distributed-order fractional optimal control problem. Section 4 develops the proposed PI-CSNN framework, including the neural-spectral representation, collocation discretization, distributed-order operator evaluation, residual formulation, physics-informed loss construction, optimization strategy, and approximation properties. Section 5 presents the numerical results and discusses the accuracy, convergence, and performance of the proposed method. Finally, Section 6 concludes the paper.

2. Preliminaries

In this section, we introduce the fundamental concepts of distributed-order fractional calculus together with the principal properties of Chebyshev polynomials that form the mathematical foundation of the proposed PI-CSNN framework.

2.1. Distributed-order fractional calculus

Let $\Omega = [0, T]$ be a finite time interval and let $\Gamma(\cdot)$ denote the Gamma function defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt, \qquad z > 0. \tag{2.1}$$

Distributed-order fractional operators extend classical fractional operators by incorporating a continuum of fractional orders and therefore provide a flexible framework for describing multiscale memory and hereditary effects. We first recall the basic fractional operators required in the subsequent analysis. For a sufficiently smooth function $u(t)$, the left-sided Riemann-Liouville fractional integral of order $\alpha > 0$ is defined by

$$\mathcal{I}_0^{\alpha} u(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} u(\tau)\, d\tau, \qquad t > 0. \tag{2.2}$$

For $\alpha \in (0,1)$, the left-sided Caputo fractional derivative is given by

$${}^{C}\mathcal{D}_0^{\alpha} u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{u'(\tau)}{(t-\tau)^{\alpha}}\, d\tau. \tag{2.3}$$

The Caputo derivative is particularly suitable for physical applications because it admits classical initial conditions expressed in terms of integer-order derivatives. The right-sided Riemann-Liouville fractional integral of order $\beta > 0$ is defined by

$${}_t\mathcal{I}_T^{\beta} u(t) = \frac{1}{\Gamma(\beta)} \int_t^T (\tau - t)^{\beta-1} u(\tau)\, d\tau. \tag{2.4}$$

Accordingly, for $\alpha \in (0,1)$, the right-sided Riemann-Liouville fractional derivative is defined by

$${}^{RL}\mathcal{D}_T^{\alpha} u(t) = -\frac{d}{dt}\left({}_t\mathcal{I}_T^{1-\alpha} u(t)\right) = -\frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_t^T \frac{u(\tau)}{(\tau - t)^{\alpha}}\, d\tau. \tag{2.5}$$

To describe systems involving multiple memory scales, let $\omega(\alpha)$ be a nonnegative weight function satisfying

$$\omega(\alpha) \geq 0, \qquad \int_0^1 \omega(\alpha)\, d\alpha < \infty. \tag{2.6}$$

The distributed-order Caputo derivative is defined by

$${}^{C}\mathfrak{D}_0^{(\omega)} u(t) = \int_0^1 \omega(\alpha)\, {}^{C}\mathcal{D}_0^{\alpha} u(t)\, d\alpha. \tag{2.7}$$

Likewise, the distributed-order right-sided Riemann-Liouville derivative is defined by

$${}^{RL}\mathfrak{D}_T^{(\omega)} u(t) = \int_0^1 \omega(\alpha)\, {}^{RL}\mathcal{D}_T^{\alpha} u(t)\, d\alpha. \tag{2.8}$$

For sufficiently regular functions u and v, the fractional integration-by-parts identity for each $\alpha \in (0,1)$ is

$$\int_0^T v(t)\, {}^{C}\mathcal{D}_0^{\alpha} u(t)\, dt = \int_0^T u(t)\, {}^{RL}\mathcal{D}_T^{\alpha} v(t)\, dt + \Big[u(t)\, {}_t\mathcal{I}_T^{1-\alpha} v(t)\Big]_0^T. \tag{2.9}$$

Integrating this identity with respect to the fractional-order variable yields the distributed-order integration-by-parts formula

$$\int_0^T v(t)\, {}^{C}\mathfrak{D}_0^{(\omega)} u(t)\, dt = \int_0^T u(t)\, {}^{RL}\mathfrak{D}_T^{(\omega)} v(t)\, dt + \int_0^1 \omega(\alpha) \Big[u(t)\, {}_t\mathcal{I}_T^{1-\alpha} v(t)\Big]_0^T\, d\alpha. \tag{2.10}$$

Hence, the boundary contribution is explicitly given by

$$\mathcal{B}_{\omega}(u, v) = \int_0^1 \omega(\alpha) \Big[u(t)\, {}_t\mathcal{I}_T^{1-\alpha} v(t)\Big]_0^T\, d\alpha. \tag{2.11}$$

This explicit form is used in the variational derivation of the adjoint equation and the associated transversality condition. For numerical implementation, the distributed-order operators introduced above are approximated by Gauss-Legendre quadrature over the fractional-order variable. The corresponding numerical discretization and implementation details are presented in Section 4.

2.2. Chebyshev polynomials

Chebyshev polynomials provide the spectral basis employed in the proposed PI-CSNN framework because of their orthogonality, excellent interpolation properties, and rapid convergence for sufficiently smooth functions. The Chebyshev polynomials of the first kind are defined on $[-1, 1]$ by

$$\mathbb{T}_0(x) = 1, \quad \mathbb{T}_1(x) = x, \tag{2.12}$$

and satisfy the recurrence relation

$$\mathbb{T}_{n+1}(x) = 2x\,\mathbb{T}_n(x) - \mathbb{T}_{n-1}(x), \quad n \geq 1. \tag{2.13}$$

Equivalently, they admit the trigonometric representation

$$\mathbb{T}_n(x) = \cos\left(n \arccos x\right), \quad x \in [-1, 1]. \tag{2.14}$$

The Chebyshev polynomials are orthogonal with respect to the weight function

$$w(x) = \frac{1}{\sqrt{1-x^2}}, \qquad x \in (-1,\ 1), \tag{2.15}$$

namely,

$$\int_{-1}^{1} \frac{\mathbb{T}_m(x)\mathbb{T}_n(x)}{\sqrt{1-x^2}}\, dx = 0, \quad m \neq n. \tag{2.16}$$

These properties make Chebyshev polynomials well-suited for global spectral approximation. In the proposed PI-CSNN framework, they are used to construct the boundary-adapted spectral trial spaces introduced in Section 4, while the corresponding spectral coefficients are generated by the neural network and optimized through the physics-informed residual formulation.

3. Mathematical formulation of the distributed-order fractional optimal control problem

Let $\Omega = [0, T]$ denote the temporal domain, and let $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ represent the state and control variables, respectively. The objective is to determine an admissible control that minimizes the performance index

$$\min_{u \in \mathcal{U}_{ad}} \mathcal{J}(x, u) = \int_0^T \mathcal{L}(t, x(t), u(t))\, dt, \tag{3.1}$$

subject to the distributed-order fractional state equation

$${}^{C}\mathfrak{D}_0^{(\omega)} x(t) = \mathcal{F}(t, x(t), u(t)), \quad 0 < t \leq T, \tag{3.2}$$

with the initial condition

$$x(0) = x_0. \tag{3.3}$$

Here, $\mathcal{L}(t, x, u)$ denotes the running cost functional, $\mathcal{F}(t, x, u)$ is the distributed-order fractional state operator, and ${}^{C}\mathfrak{D}_0^{(\omega)}$ denotes the distributed-order Caputo derivative associated with the weight function $\omega(\alpha)$. The admissible control set is defined by

$$\mathcal{U}_{ad} = \left\{ u \in L^2(0, T; \mathbb{R}^m) : u_{\min}(t) \leq u(t) \leq u_{\max}(t) \text{ a.e. in } [0, T] \right\}. \tag{3.4}$$

We assume that $\mathcal{U}_{ad}$ is nonempty, closed, bounded, and convex, that $\mathcal{L}(t, x, u)$ is continuous and convex with respect to the control variable, and that $\mathcal{F}(t, x, u)$ satisfies the continuity and Lipschitz conditions required for the well-posedness of the state equation. The weight function $\omega(\alpha)$ is assumed to be nonnegative and integrable on $[0, 1]$. Together with the standard regularity and coercivity assumptions, these conditions ensure the well-posedness of the state system and provide the standard setting for the existence of an optimal solution.

Let $\psi(t) \in \mathbb{R}^n$ denote the adjoint variable. The Hamiltonian associated with the optimal control problem is defined by

$$\mathcal{H}(t, x, u, \psi) = \mathcal{L}(t, x, u) + \psi^T(t)\mathcal{F}(t, x, u). \tag{3.5}$$

The corresponding augmented functional is

$$\widetilde{\mathcal{J}} = \int_0^T [\mathcal{H}(t, x, u, \psi) - \psi^T(t)\, {}^{C}\mathfrak{D}_0^{(\omega)} x(t)]\, dt. \tag{3.6}$$

Taking the first variation with respect to the state variable and applying the distributed-order fractional integration-by-parts formula yields

$$\delta_x \widetilde{\mathcal{J}} = \int_0^T \delta x^T(t)[\frac{\partial \mathcal{H}}{\partial x} - {}^{RL}\mathfrak{D}_T^{(\omega)} \psi(t)]\, dt + \mathcal{B}(\delta x, \psi), \tag{3.7}$$

where $\mathcal{B}(\delta x, \psi)$ denotes the boundary contribution arising from the distributed-order fractional integration-by-parts formula. Since the initial state is prescribed, $\delta x(0) = 0$. For a free terminal state without a terminal cost, the transversality condition becomes

$$\psi(T) = 0. \tag{3.8}$$

Consequently, stationarity of the augmented functional with respect to the independent variations yields the distributed-order fractional optimality system

$${}^{C}\mathfrak{D}_0^{(\omega)} x(t) = \frac{\partial \mathcal{H}}{\partial \psi}, \quad x(0) = x_0, \tag{3.9a}$$

$${}^{RL}\mathfrak{D}_T^{(\omega)} \psi(t) = \frac{\partial \mathcal{H}}{\partial x}, \quad \psi(T) = 0. \tag{3.9b}$$

For box-constrained controls, the first-order necessary optimality condition is expressed by the variational inequality

$$\int_0^T \left(\frac{\partial \mathcal{H}}{\partial u}(t) \right)^T (v(t) - u(t))\, dt \geq 0, \quad \forall\, v \in \mathcal{U}_{ad}. \tag{3.10}$$

For an unconstrained interior optimal control, this reduces to the stationarity condition

$$\frac{\partial \mathcal{H}}{\partial u} = 0. \tag{3.11}$$

Equations (3.9a) and (3.9b), together with the appropriate control optimality condition, constitute the continuous distributed-order fractional optimality system that forms the mathematical foundation of the proposed PI-CSNN framework.

4. Physics-informed Chebyshev spectral neural network methodology

In this section, we develop the proposed PI-CSNN framework for solving the distributed-order fractional optimal control problems under consideration. The methodology combines neural feature learning, Chebyshev spectral approximation, and physics-informed residual minimization within a unified computational framework for determining the state, adjoint, and control variables directly from the governing optimality system without requiring labeled solution data. Rather than optimizing the spectral coefficients independently, the PI-CSNN framework learns a shared latent representation of the temporal coordinate from which the global Chebyshev coefficient vectors are generated, while the governing optimality system is enforced through physics-informed residual minimization. For clarity, the methodology is presented for the scalar case, whereas the extension to coupled vector-valued systems follows componentwise. The distributed-order fractional operators are approximated by Gauss-Legendre quadrature in the fractional-order variable, while the temporal dependence of the unknown variables is represented through boundary-adapted shifted Chebyshev expansions. The prescribed initial condition for the state and terminal condition for the adjoint variable are incorporated exactly into the trial representations, thereby eliminating the need for corresponding penalty terms in the loss functional. The governing state, adjoint, and optimality equations are enforced through physics-informed residuals evaluated at temporal collocation points. When an explicit control law is unavailable, the control is approximated independently and the associated optimality condition is incorporated into the residual formulation. For box-constrained control problems, the corresponding variational inequality or projection formulation may be employed.

4.1. Chebyshev spectral neural representation

Let $\mathbb{T}_j(\xi)$ denote the Chebyshev polynomial of the first kind of degree j. The physical interval $[0, T]$ is mapped onto $[-1, 1]$ through

$$\xi(t) = \frac{2t}{T} - 1, \tag{4.1}$$

and the corresponding shifted Chebyshev functions are defined by

$$\widehat{\mathbb{T}}_j(t) = \mathbb{T}_j(\xi(t)), \quad j = 0, 1, \ldots, M. \tag{4.2}$$

The normalized temporal coordinate $\xi(t)$ is supplied as a nonconstant input to a feedforward neural feature extractor

$$\mathcal{N}_{\Theta_{\mathcal{N}}} : [-1, 1] \longrightarrow \mathbb{R}^r, \tag{4.3}$$

with hidden-layer recursion

$$h^{(0)}(\xi) = \xi, \tag{4.4}$$

$$h^{(\ell)}(\xi) = \sigma\left(\mathbf{W}^{(\ell)} h^{(\ell-1)}(\xi) + \mathbf{d}^{(\ell)}\right), \quad \ell = 1, \ldots, L, \tag{4.5}$$

where $W^{(\ell)}$ and $d^{(\ell)}$ denote the trainable weight matrices and bias vectors, respectively, and σ is the activation function. Throughout this work, the hyperbolic tangent activation function is employed. At the temporal nodes $\{t_i\}_{i=1}^{N_c}$, the network produces the latent feature vectors

$$h_i = h^{(L)}(\xi(t_i)) \in \mathbb{R}^r, \quad i = 1, \ldots, N_c. \tag{4.6}$$

Since the state, adjoint, and control variables are represented through global Chebyshev expansions, the nodewise latent features are aggregated into the pooled representation

$$\overline{h} = \sum_{i=1}^{N_c} \rho_i h_i, \quad \rho_i \geq 0, \quad \sum_{i=1}^{N_c} \rho_i = 1, \tag{4.7}$$

where, when temporal quadrature weights $\omega_i^{(t)}$ are available,

$$\rho_i = \frac{\omega_i^{(t)}}{\sum_{\ell=1}^{N_c} \omega_\ell^{(t)}}. \tag{4.8}$$

The pooled feature vector is subsequently mapped to the spectral coefficient vectors through trainable linear output heads:

$$\begin{aligned} a(\Theta) &= W_a \overline{h} + d_a, \\ b(\Theta) &= W_b \overline{h} + d_b, \\ c(\Theta) &= W_c \overline{h} + d_c, \end{aligned} \tag{4.9}$$

where

$$a(\Theta),\ b(\Theta),\ c(\Theta) \in \mathbb{R}^{M+1}. \tag{4.10}$$

Their components are denoted by $a_j(\Theta)$, $b_j(\Theta)$, and $c_j(\Theta)$, $j = 0, \ldots, M$. To satisfy the prescribed boundary conditions exactly, the state and adjoint variables are approximated by the boundary-adapted trial functions

$$x_M(t;\Theta) = x_0 + t \sum_{j=0}^{M} a_j(\Theta) \widehat{\mathbb{T}}_j(t) \tag{4.11}$$

and

$$\psi_M(t;\Theta) = (T - t) \sum_{j=0}^{M} b_j(\Theta) \widehat{\mathbb{T}}_j(t), \tag{4.12}$$

which satisfy

$$x_M(0;\Theta) = x_0, \quad \psi_M(T;\Theta) = 0, \tag{4.13}$$

for every admissible parameter set Θ. The endpoint values $x_M(T;\Theta)$ and $\psi_M(0;\Theta)$ remain unconstrained. When the control is treated as an independent unknown, it is represented by

$$u_M(t;\Theta) = \sum_{j=0}^{M} c_j(\Theta) \widehat{\mathbb{T}}_j(t), \tag{4.14}$$

so that no artificial endpoint condition is imposed. For problems in which the stationarity condition yields an explicit control law, the independent control output head is omitted and the control is reconstructed directly from the state and adjoint approximations. The complete trainable parameter set is written as

$$\Theta = \Theta_{\mathcal{N}} \cup \Theta_{\text{out}}, \tag{4.15}$$

where

$$\Theta_{\mathcal{N}} = \{W^{(\ell)},\ d^{(\ell)}\}_{\ell=1}^{L} \tag{4.16}$$

contains the hidden-layer parameters and Θ_{out} denotes the parameters of the active output heads. The proposed PI-CSNN framework employs the neural network as a feature-learning component rather than merely as an alternative parameterization of the Chebyshev spectral coefficients. The trainable parameter set Θ defines a nonlinear mapping from the normalized temporal coordinate to a latent feature representation, from which the global spectral coefficient vectors are generated through the output heads. In this way, the neural network provides a common nonlinear parameterization for all spectral coefficient vectors while preserving the global spectral structure of the underlying Chebyshev approximation. Consequently, the coefficient vectors are not optimized independently but are coupled through a shared latent representation learned by the neural network. The trainable parameters are determined by minimizing the physics-informed residual objective associated with the distributed-order optimality system. During training, the neural feature extractor learns a latent representation of the temporal coordinate, while the output heads transform this representation into the global Chebyshev coefficient vectors defining the state, adjoint, and, when required, control approximations. Consequently, the optimization is performed in the neural parameter space Θ, whereas the governing equations are enforced through the resulting spectral representations. Accordingly, the PI-CSNN optimization seeks the neural parameter set Θ whose induced spectral approximations satisfy the distributed-order optimality system in the physics-informed least-squares sense.

Direct optimization of the spectral coefficients remains an efficient strategy for classical problem-specific spectral collocation methods. The objective of the proposed PI-CSNN framework is not to replace such approaches, but to provide a unified trainable architecture that integrates neural feature learning, global Chebyshev spectral approximation, distributed-order operator evaluation, and physics-informed residual minimization within a single optimization framework. Rather than optimizing the spectral coefficients independently, the proposed approach learns a shared latent representation from which all coefficient vectors are generated, enabling a common optimization strategy for state, adjoint, and control variables. This formulation naturally extends to distributed-order, nonlinear, inverse, and parameter-identification problems without requiring a reformulation of the underlying optimization procedure.

4.2. Chebyshev collocation and residual discretization

The spectral degree M determines the dimension of the global Chebyshev approximation, whereas the governing equations are enforced at an independently selected set of temporal collocation points. This separation between approximation and residual evaluation naturally permits over-collocation, thereby providing additional flexibility for resolving the physics-informed residuals without increasing the number of retained spectral modes.

Let $\{t_i, \omega_i^{(t)}\}_{i=1}^{N_c}$ denote the N_c-point Gauss-Legendre quadrature rule mapped onto the physical interval $[0, T]$, where t_i and $\omega_i^{(t)}$ are the corresponding temporal nodes and quadrature weights. The associated reference coordinates are

$$\xi_i = \frac{2t_i}{T} - 1, \quad i = 1, \dots, N_c. \tag{4.17}$$

These normalized coordinates serve as inputs to the neural feature extractor, whereas the corresponding physical nodes are used to evaluate the governing state, adjoint, and optimality residuals. Consequently, M determines the dimension of the global spectral approximation, whereas N_c controls the temporal resolution used for residual enforcement. In general, $N_c \neq M + 1$, and choosing $N_c > M + 1$ yields an over-collocated discretization. We define the shifted Chebyshev evaluation matrix

$$\mathbf{T} = [\widehat{\mathbb{T}}_j(t_i)]_{i=1,\dots,N_c,\ j=0,\dots,M}, \tag{4.18}$$

together with the boundary-adapted basis matrices

$$\mathbf{B}_x = [t_i \widehat{\mathbb{T}}_j(t_i)]_{i=1,\dots,N_c,\ j=0,\dots,M} \tag{4.19}$$

and

$$\mathbf{B}_\psi = [(T - t_i)\widehat{\mathbb{T}}_j(t_i)]_{i=1,\dots,N_c,\ j=0,\dots,M}. \tag{4.20}$$

Once the PI-CSNN generates the spectral coefficient vectors $a(\Theta)$ and $b(\Theta)$, the corresponding nodal state and adjoint approximations are obtained as

$$\mathbf{x}(\Theta) = x_0\mathbf{1} + \mathbf{B}_x a(\Theta), \quad \boldsymbol{\psi}(\Theta) = \mathbf{B}_\psi b(\Theta). \tag{4.21}$$

When the control is represented independently, its nodal values are evaluated by

$$\mathbf{u}(\Theta) = \mathbf{T}c(\Theta). \tag{4.22}$$

Thus, the collocation procedure evaluates the neural-generated global spectral approximations directly at the temporal nodes to construct the physics-informed residuals. Unlike a conventional Chebyshev collocation method, the resulting collocation equations are not solved directly for the spectral coefficient vectors. Instead, they define the residuals used to optimize the neural parameter set Θ, where its induced spectral approximations satisfy the governing optimality system. The temporal quadrature weights are assembled into the diagonal matrix

$$\mathbf{W}_t = \operatorname{diag}(\omega_1^{(t)}, \omega_2^{(t)}, \dots, \omega_{N_c}^{(t)}), \tag{4.23}$$

which defines the weighted discrete norm

$$\|\mathbf{R}\|_{\mathbf{W}_t}^2 = \mathbf{R}^T\mathbf{W}_t\mathbf{R} = \sum_{i=1}^{N_c} \omega_i^{(t)} |R(t_i)|^2, \tag{4.24}$$

for any residual vector $\mathbf{R}$ evaluated at the temporal nodes. This weighted norm provides the discrete measure employed in the physics-informed loss functional. The temporal and fractional-order discretizations play complementary roles within the proposed framework. The shifted Chebyshev basis defines the global approximation space, whereas the temporal Gauss-Legendre nodes are used for neural feature evaluation, residual enforcement, and weighted integration. The distributed-order operators are discretized independently through Gauss-Legendre quadrature in the fractional-order variable. For Dirac-delta weight distributions, the distributed-order operators reduce exactly to the corresponding single-order fractional operators, so no quadrature approximation of the delta distribution is introduced.

4.3. Residual formulation

The distributed-order fractional optimality system is enforced through physics-informed residuals evaluated at the temporal collocation nodes $\{t_i\}_{i=1}^{N_c}$. Since the state, adjoint, and, when required, control approximations are generated by the neural-induced Chebyshev spectral coefficients, all residuals depend implicitly on the trainable parameter set Θ. The state residual is defined by

$$\mathcal{R}_x(t_i;\Theta) = {}^{C}\mathfrak{D}_0^{(\omega)} x_M(t_i;\Theta) - \mathcal{F}(t_i, x_M(t_i;\Theta), u_M(t_i;\Theta)), \quad i = 1,\ldots,N_c. \tag{4.25}$$

Similarly, the adjoint residual is given by

$$\mathcal{R}_\psi(t_i;\Theta) = {}^{RL}\mathfrak{D}_T^{(\omega)} \psi_M(t_i;\Theta) - \frac{\partial \mathcal{H}}{\partial x}(t_i, x_M(t_i;\Theta), u_M(t_i;\Theta), \psi_M(t_i;\Theta)), \quad i = 1,\ldots,N_c. \tag{4.26}$$

When the control is approximated independently, the corresponding optimality residual is defined by

$$\mathcal{R}_u(t_i;\Theta) = \frac{\partial \mathcal{H}}{\partial u}(t_i, x_M(t_i;\Theta), u_M(t_i;\Theta), \psi_M(t_i;\Theta)), \quad i = 1,\ldots,N_c. \tag{4.27}$$

For compact notation, the state and adjoint residual vectors are written as

$$\mathbf{R}_x(\Theta) = [\mathcal{R}_x(t_i;\Theta)]_{i=1}^{N_c}, \quad \mathbf{R}_\psi(\Theta) = [\mathcal{R}_\psi(t_i;\Theta)]_{i=1}^{N_c}, \tag{4.28}$$

and, when the control is represented independently,

$$\mathbf{R}_u(\Theta) = [\mathcal{R}_u(t_i;\Theta)]_{i=1}^{N_c}. \tag{4.29}$$

Since the prescribed state initial condition and adjoint terminal condition are incorporated exactly into the boundary-adapted trial representations, no additional residual terms are required for these conditions. Any remaining boundary, terminal, or inequality constraints that cannot be embedded directly into the trial functions may be enforced through corresponding residual contributions. Accordingly, the residual vectors provide a quantitative measure of the extent to which the neural-generated spectral approximations satisfy the distributed-order optimality system at the collocation nodes. Minimizing these residuals yields physics-informed approximations of the state, adjoint, and, when required, control variables without requiring labeled solution data.

4.4. Approximation and evaluation of distributed-order fractional derivatives

The distributed-order fractional operators appearing in the governing optimality system are evaluated numerically using Gauss-Legendre quadrature over the fractional-order interval.

Let $\{\alpha_k, \omega_k^{(\alpha)}\}_{k=1}^{K}$ denote the corresponding K-point quadrature rule on $[\alpha_{\min}, \alpha_{\max}]$. For a prescribed smooth weight function $\omega(\alpha)$, the effective quadrature coefficients are defined by

$$\vartheta_k = \omega_k^{(\alpha)} \omega(\alpha_k), \quad k = 1,\ldots,K. \tag{4.30}$$

The distributed-order Caputo and Riemann-Liouville operators are then approximated by

$${}^{C}\mathfrak{D}_0^{(\omega)} f(t) \approx \sum_{k=1}^{K} \vartheta_k \, {}^{C}\mathcal{D}_0^{\alpha_k} f(t), \quad {}^{RL}\mathfrak{D}_T^{(\omega)} f(t) \approx \sum_{k=1}^{K} \vartheta_k \, {}^{RL}\mathcal{D}_T^{\alpha_k} f(t). \tag{4.31}$$

When a continuous density is normalized numerically on a truncated fractional-order interval, the effective coefficients become

$$\widehat{\vartheta}_k = \frac{\omega_k^{(\alpha)}\omega(\alpha_k)}{\sum_{\ell=1}^{K}\omega_\ell^{(\alpha)}\omega(\alpha_\ell)}, \quad k = 1, \ldots, K. \tag{4.32}$$

For a Dirac-delta weight distribution concentrated at $\alpha = \alpha^*$, the distributed-order operator reduces exactly to the corresponding single-order fractional operator, and no quadrature approximation of the delta distribution is introduced. Since the PI-CSNN approximations are expressed as finite boundary-adapted shifted Chebyshev expansions, the required fractional derivatives are evaluated termwise after expressing the basis functions in polynomial form. The shifted Chebyshev functions satisfy

$$\widehat{\mathbb{T}}_j(t) = \mathbb{T}_j(\frac{2t}{T} - 1) = \sum_{m=0}^{j} \gamma_{j,m}^{(T)} t^m, \tag{4.33}$$

where $\gamma_{j,m}^{(T)}$ are determined recursively from the Chebyshev polynomials and the interval transformation.

For $n-1 < \alpha < n$, the left-sided Caputo derivative of a monomial is given by

$${}^{C}\mathcal{D}_0^\alpha t^m = \begin{cases} 0, & m = 0, 1, \ldots, n-1, \\ \dfrac{\Gamma(m+1)}{\Gamma(m-\alpha+1)} t^{m-\alpha}, & m \geq n. \end{cases} \tag{4.34}$$

Consequently,

$${}^{C}\mathcal{D}_0^\alpha x_M(t;\Theta) = \sum_{j=0}^{M} a_j(\Theta)\, {}^{C}\mathcal{D}_0^\alpha[t\,\widehat{\mathbb{T}}_j(t)], \tag{4.35}$$

since the prescribed initial value x_0 has a zero Caputo derivative. For the adjoint approximation, we introduce the reflected coordinate

$$s = T - t, \tag{4.36}$$

so that

$$\psi_M(T-s;\Theta) = \sum_{m=1}^{M+1} \eta_m(\Theta)\, s^m, \tag{4.37}$$

where the coefficients $\eta_m(\Theta)$ depend linearly on $b(\Theta)$. Using

$${}^{RL}\mathcal{D}_T^\alpha (T-t)^m = \frac{\Gamma(m+1)}{\Gamma(m-\alpha+1)}\,(T-t)^{m-\alpha}, \tag{4.38}$$

the right-sided Riemann-Liouville derivative is evaluated directly as

$${}^{RL}\mathcal{D}_T^\alpha \psi_M(t;\Theta) = \sum_{m=1}^{M+1} \eta_m(\Theta)\, \frac{\Gamma(m+1)}{\Gamma(m-\alpha+1)}(T-t)^{m-\alpha}. \tag{4.39}$$

This avoids replacing the right-sided operator by a reflected left-sided Caputo derivative.

To improve computational efficiency during the repeated residual evaluations required by the optimization process, the fractional actions of the boundary-adapted basis functions are precomputed at the temporal collocation nodes. We define

$$\mathbf{D}_C^{(\alpha_k)} = \left[{}^{C}\mathcal{D}_0^{\alpha_k}\left(t\widehat{\mathbb{T}}_j(t)\right)\Big|_{t=t_i}\right]_{i=1,\dots,N_c,\ j=0,\dots,M} \tag{4.40}$$

and

$$\mathbf{D}_{RL}^{(\alpha_k)} = \left[{}^{RL}\mathcal{D}_T^{\alpha_k}\left((T-t)\widehat{\mathbb{T}}_j(t)\right)\Big|_{t=t_i}\right]_{i=1,\dots,N_c,\ j=0,\dots,M}. \tag{4.41}$$

For continuous distributed-order densities, the corresponding operator matrices are assembled as

$$\mathbf{D}_C^{(\omega)} = \sum_{k=1}^{K} \vartheta_k\, \mathbf{D}_C^{(\alpha_k)}, \quad \mathbf{D}_{RL}^{(\omega)} = \sum_{k=1}^{K} \vartheta_k\, \mathbf{D}_{RL}^{(\alpha_k)}. \tag{4.42}$$

When normalized coefficients are employed, $\widehat{\vartheta}_k$ replaces ϑ_k in (4.42). Accordingly,

$$\mathbf{d}_x^{(\omega)}(\Theta) = \mathbf{D}_C^{(\omega)} a(\Theta), \quad \mathbf{d}_\psi^{(\omega)}(\Theta) = \mathbf{D}_{RL}^{(\omega)} b(\Theta). \tag{4.43}$$

The distributed-order operator matrices depend only on the selected spectral basis and the fractional-order quadrature rule, and are therefore assembled once during an offline preprocessing stage. Throughout the subsequent optimization, only the neural parameter set Θ is updated, while the precomputed operator matrices are reused to evaluate the distributed-order derivatives through efficient matrix-vector products. This offline-online decomposition substantially reduces the computational overhead associated with repeated residual evaluations during training. The parameters M and K play complementary numerical roles. The spectral degree M determines the approximation space, whereas K controls the resolution of the distributed-order integral. Consequently, the overall approximation accuracy depends jointly on the spectral truncation, distributed-order quadrature, temporal residual discretization, and the subsequent optimization process.

4.5. *Physics-informed loss functional*

The trainable parameter set Θ is determined by minimizing a physics-informed objective constructed from the residuals of the distributed-order optimality system. Using the temporal collocation nodes and the associated quadrature weights, the state and adjoint loss terms are defined by

$$\mathcal{L}_{\text{state}}(\Theta) = \sum_{i=1}^{N_c} \omega_i^{(t)} \left|\mathcal{R}_x(t_i;\Theta)\right|^2 \tag{4.44}$$

and

$$\mathcal{L}_{\text{adjoint}}(\Theta) = \sum_{i=1}^{N_c} \omega_i^{(t)} \left|\mathcal{R}_\psi(t_i;\Theta)\right|^2. \tag{4.45}$$

When the control is represented independently, the corresponding optimality loss is

$$\mathcal{L}_{\text{optimal}}(\Theta) = \sum_{i=1}^{N_c} \omega_i^{(t)} \left|\mathcal{R}_u(t_i;\Theta)\right|^2. \tag{4.46}$$

The total physics-informed objective is therefore

$$\mathcal{L}_{\text{total}}(\Theta) = \lambda_x \mathcal{L}_{\text{state}}(\Theta) + \lambda_\psi \mathcal{L}_{\text{adjoint}}(\Theta) + \lambda_u \mathcal{L}_{\text{optimal}}(\Theta), \tag{4.47}$$

where $\lambda_x, \lambda_\psi, \lambda_u \geq 0$ are nonnegative residual weights. Unless otherwise stated, equal weights are employed throughout the numerical experiments. When the control is recovered directly from the optimality condition, $\mathcal{L}_{\text{optimal}}$ is omitted. Additional constraints that cannot be incorporated directly into the trial representations may be included through corresponding residual contributions. The PI-CSNN training problem is formulated as

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}_{\text{total}}(\Theta). \tag{4.48}$$

Since the distributed-order operator matrices are precomputed from the spectral basis and quadrature rules, the objective function remains differentiable with respect to the trainable parameter set Θ and can therefore be minimized efficiently using gradient-based optimization.

4.6. Optimization strategy and computational complexity

The trainable parameter set Θ is determined using a two-stage gradient-based optimization strategy. The physics-informed objective is first minimized using the Adam optimizer to obtain a suitable initial estimate, after which the best iterate is employed to initialize a quasi-Newton refinement. Gradients are computed by backpropagation through the neural feature extractor and the spectral output heads, whereas the distributed-order fractional operators associated with the boundary-adapted basis functions are assembled only once before training and subsequently reused throughout the optimization process.

Let M denote the Chebyshev spectral degree, K the number of fractional-order quadrature nodes, N_c the number of temporal collocation points, and P the number of trainable parameters. The offline assembly of the distributed-order fractional operator matrices requires approximately $\mathcal{O}(K\, N_c\, M)$ operations. Once these matrices have been assembled, a representative loss-and-gradient evaluation requires

$$\mathcal{O}\left(N_c \sum_{\ell=1}^{L} n_{\ell-1} n_\ell + N_c M + P\right), \tag{4.49}$$

where $n_0, n_1, \ldots, n_L$ denote the widths of the neural network layers. The first term accounts for forward and backward propagation through the neural feature extractor, the second corresponds to spectral reconstruction together with residual evaluation, and the final term represents the dominant parameter-dependent gradient computation. The overall computational cost additionally depends on the number of optimization iterations and the active residual components included in the physics-informed objective. Consequently, the proposed PI-CSNN framework naturally separates into an offline preprocessing stage and an online optimization stage. During preprocessing, the distributed-order fractional operator matrices are assembled once from the spectral basis and quadrature rule. During training, only the neural parameter set Θ is updated, while the precomputed operator matrices are reused throughout every optimization iteration. This offline-online decomposition avoids repeated assembly of the distributed-order operators and provides an efficient implementation of the proposed neural-spectral framework. The overall computational workflow of the proposed PI-CSNN framework is summarized in Figure 1, illustrating the interaction between neural feature learning, Chebyshev spectral approximation, distributed-order operator evaluation, physics-informed residual construction, and gradient-based optimization.

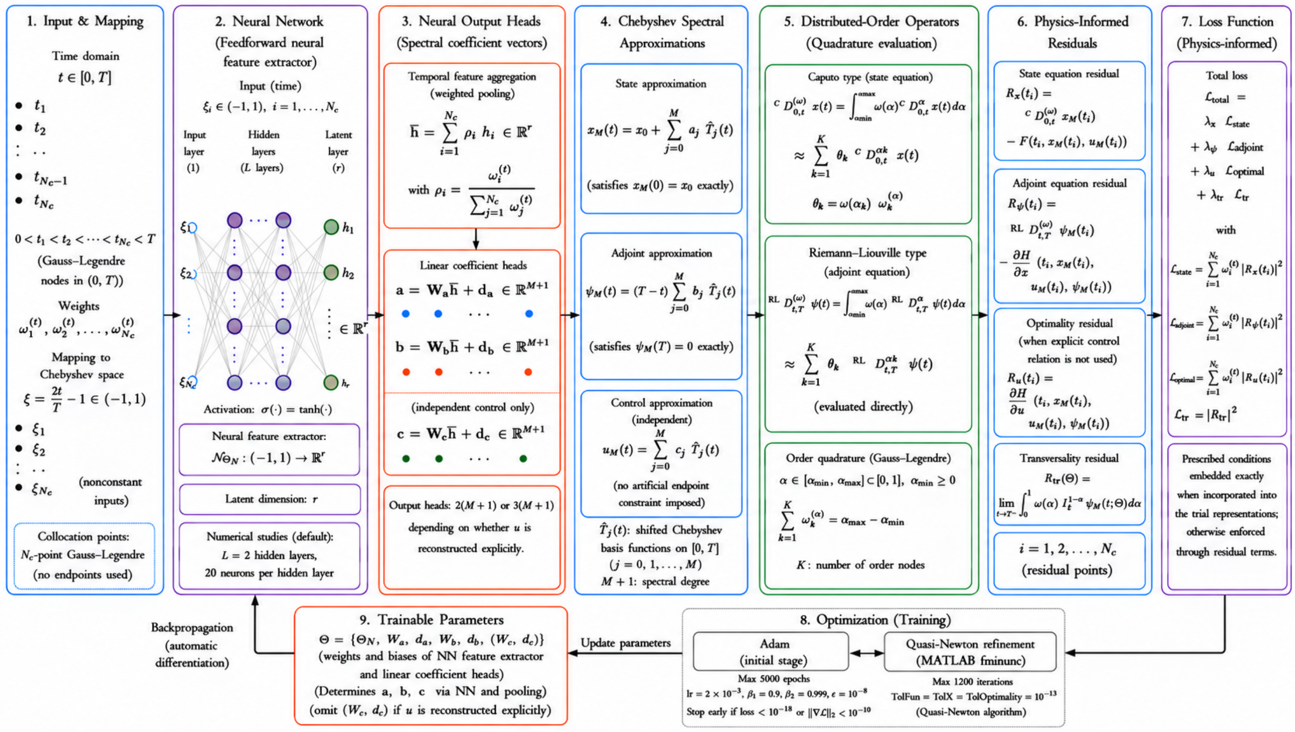

Figure 1. Overall computational workflow of the proposed PI-CSNN framework.

4.7. Approximation and convergence properties

The approximation properties of the proposed PI-CSNN framework are inherited from the underlying boundary-adapted Chebyshev trial spaces, whereas the neural network provides a trainable parameterization of the corresponding spectral coefficients. The results presented in this subsection therefore characterize the approximation capability of the admissible trial spaces rather than a convergence proof for the complete physics-informed optimization algorithm. To describe these approximation properties, let

$$\mathbb{P}_M = \operatorname{span}\{\widehat{\mathbb{T}}_0, \widehat{\mathbb{T}}_1, \ldots, \widehat{\mathbb{T}}_M\} \tag{4.50}$$

and define

$$\mathcal{V}_M^x = \{x_0 + t\, p_M(t) : p_M \in \mathbb{P}_M\}, \quad \mathcal{V}_M^\psi = \{(T - t)\, r_M(t) : r_M \in \mathbb{P}_M\}. \tag{4.51}$$

For an independently approximated control,

$$\mathcal{V}_M^u = \mathbb{P}_M, \tag{4.52}$$

whereas no independent approximation space is required when the control is recovered directly from the optimality condition. The factors t and $T - t$ ensure exact satisfaction of the prescribed state initial condition and adjoint terminal condition, respectively. Assume that the boundary-compatible transformed functions

$$\widetilde{x}(t) = \frac{x(t) - x_0}{t}, \quad \widetilde{\psi}(t) = \frac{\psi(t)}{T - t}, \tag{4.53}$$

with endpoint values defined by continuous extension, belong to $H^m(\Omega)$ for some $m \geq 1$. If the control is approximated independently, assume also that $u \in H^m(\Omega)$. Standard Chebyshev approximation theory then gives

$$\inf_{p_M \in \mathbb{P}_M} \|\widetilde{x} - p_M\|_{L^2(\Omega)} \leq C_x \, M^{-m} \|\widetilde{x}\|_{H^m(\Omega)}, \tag{4.54}$$

$$\inf_{r_M \in \mathbb{P}_M} \|\widetilde{\psi} - r_M\|_{L^2(\Omega)} \leq C_\psi \, M^{-m} \|\widetilde{\psi}\|_{H^m(\Omega)}, \tag{4.55}$$

$$\inf_{s_M \in \mathbb{P}_M} \|u - s_M\|_{L^2(\Omega)} \leq C_u \, M^{-m} \|u\|_{H^m(\Omega)}. \tag{4.56}$$

Since the factors t and $T - t$ remain bounded on Ω, the boundary-adapted trial spaces satisfy

$$\inf_{v_M \in \mathcal{V}_M^x} \|x - v_M\|_{L^2(\Omega)} \leq C M^{-m}, \tag{4.57}$$

$$\inf_{y_M \in \mathcal{V}_M^\psi} \|\psi - y_M\|_{L^2(\Omega)} \leq C M^{-m}, \tag{4.58}$$

with an analogous estimate for the independently approximated control. The constants are independent of M, although they may depend on the solution regularity and the length of the time interval. The preceding estimates describe the best-approximation properties of the boundary-adapted Chebyshev trial spaces and, consequently, the maximum approximation accuracy attainable by the PI-CSNN representation for a given spectral degree. They do not constitute a convergence proof for the complete PI-CSNN optimization procedure, whose overall accuracy also depends on the distributed-order quadrature, temporal residual discretization, and the optimization process. Accordingly, the total numerical error may be decomposed schematically as

$$E_{\text{total}} \lesssim E_{\text{spec}} + E_{\text{ord}} + E_{\text{time}} + E_{\text{opt}}, \tag{4.59}$$

where E_{spec} denotes the spectral approximation error, E_{ord} the distributed-order quadrature error, E_{time} the temporal residual-discretization error, and E_{opt} the optimization error. For sufficiently smooth dependence of the distributed-order integrand on the fractional-order variable, Gauss-Legendre quadrature converges rapidly as the number of quadrature nodes K increases. Furthermore, if the boundary-compatible transformed functions are analytic in a complex neighborhood of Ω, classical Chebyshev approximation theory yields the exponential estimate

$$E_{\text{spec}} = O\left(e^{-\sigma M}\right), \quad \sigma > 0,$$

whereas solutions of finite Sobolev regularity exhibit the algebraic convergence described by Eqs (4.54)–(4.58). Consequently, the proposed PI-CSNN framework is expected to inherit the approximation properties of the underlying Chebyshev trial spaces whenever the exact solution possesses sufficient regularity, the distributed-order quadrature and temporal collocation are adequately resolved, and the optimization process produces sufficiently small physics-informed residuals. Under these conditions, increasing the approximation parameters M, K, and N_c systematically reduces the corresponding approximation components until the remaining error is dominated by optimization accuracy or machine precision.

5. Numerical results and discussions

In this section, numerical experiments are presented to evaluate the accuracy, convergence behavior, robustness, and computational performance of the proposed PI-CSNN framework. The examples are designed to assess the approximation capability of the proposed methodology for benchmark, time-invariant, and time-varying distributed-order fractional optimal control problems. Throughout all experiments, the normalized temporal coordinate is supplied to the neural feature extractor, whose latent representation is mapped to the global Chebyshev spectral coefficient vectors defining the state, adjoint, and, when required, control approximations. The trainable parameter set is determined exclusively by minimizing the physics-informed residual objective, and no exact, benchmark, or reference solution values are incorporated into the training process. Unless otherwise stated, all numerical experiments employ the same neural architecture and optimization settings to provide a consistent assessment of the proposed methodology. The PI-CSNN consists of two hidden layers with 20 neurons per layer and uses the hyperbolic tangent activation function, Xavier-type weight initialization, double-precision arithmetic, and a fixed random seed equal to 1 to ensure reproducibility. The default spectral discretization uses polynomial degree $M = 15$ together with $N_c = 180$ temporal Gauss-Legendre collocation points. For continuous distributed-order densities, the fractional-order integrals are approximated using $K = 40$ Gauss-Legendre quadrature nodes, whereas Dirac-delta weight distributions are evaluated directly at the prescribed fractional order without introducing quadrature approximation.

The trainable neural parameter set is optimized using a two-stage gradient-based strategy. The Adam optimizer is first employed for at most 5000 epochs with an initial learning rate of 2×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The Adam stage is terminated when either the physics-informed objective decreases below 10^{-18} or the Euclidean norm of its gradient falls below 10^{-10}, and the best iterate obtained during this stage is used to initialize the subsequent quasi-Newton refinement. The refinement is performed using MATLAB's `fminunc` routine with the quasi-Newton algorithm for at most 1200 iterations, using function, step, and optimality tolerances of 10^{-13}. Exact or benchmark solutions, when available, are employed exclusively after training for independent validation, error assessment, convergence studies, and comparison with existing numerical methods.

Example 1. This example evaluates the approximation accuracy and convergence behavior of the proposed PI-CSNN framework using a manufactured-solution benchmark. Since the analytical solution is available, it provides an objective assessment of the numerical accuracy under different levels of solution regularity. The analytical solution is employed exclusively for post-training validation and error evaluation and is not incorporated into the physics-informed loss or used as labeled training data. We consider the following DOFOCP:

$$\min_{w} J(q, w) = \int_0^1 \left[\left(t^{\alpha} - q(t)\right)^2 + \left(w(t) - (1 - t)^{\alpha}\right)^2 \right] dt, \tag{5.1}$$

subject to

$${}^{C}\mathfrak{D}_0^{(\omega)} q(t) = w(t) + f(t), \quad q(0) = 0. \tag{5.2}$$

The weight $\omega(\mu)$ is taken as a normalized Gaussian density on $[0, 1]$ with mean 0.5 and standard deviation 0.1, and the distributed-order integral is evaluated using the quadrature procedure described previously. The forcing function is chosen so that

$$q(t) = t^{\alpha}, \quad w(t) = (1-t)^{\alpha}, \tag{5.3}$$

which yields

$$f(t) = \int_0^1 \omega(\mu)\, \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-\mu+1)} t^{\alpha-\mu}\, d\mu - (1-t)^{\alpha}. \tag{5.4}$$

This benchmark permits direct examination of regularity-dependent convergence: larger values of α produce smoother profiles, whereas smaller values lead to weaker endpoint regularity near $t = 0$ for $q(t) = t^{\alpha}$ and near $t = 1$ for $w(t) = (1-t)^{\alpha}$. The Hamiltonian is

$$\mathcal{H} = (t^{\alpha} - q(t))^2 + (w(t) - (1-t)^{\alpha})^2 + z(t)(w(t) + f(t)). \tag{5.5}$$

According to the optimality framework described in Section 3, the corresponding necessary conditions consist of the state equation, the right-sided distributed-order adjoint equation, the stationarity condition, and the terminal transversality condition

$$z(1) = 0. \tag{5.6}$$

Accordingly, the problem-specific residuals are defined by

$$\mathcal{R}_q(t_i;\Theta) = {}^{C}\mathfrak{D}_0^{(\omega)}\, q_M(t_i;\Theta) - w_M(t_i;\Theta) - f(t_i), \tag{5.7}$$

$$\mathcal{R}_z(t_i;\Theta) = {}^{RL}\mathfrak{D}_1^{(\omega)}\, z_M(t_i;\Theta) - 2(q_M(t_i;\Theta) - t_i^{\alpha}), \tag{5.8}$$

$$\mathcal{R}_w(t_i;\Theta) = 2(w_M(t_i;\Theta) - (1-t_i)^{\alpha}) + z_M(t_i;\Theta). \tag{5.9}$$

The prescribed functions t_i^{α} and $(1-t_i)^{\alpha}$ in these residuals originate from the objective functional and therefore form part of the governing optimality system; they are not supervised solution-matching data. For this independent-control formulation, the PI-CSNN generates the coefficient vectors $a(\Theta)$–$c(\Theta)$ defining

$$q_M(t;\Theta) = t \sum_{j=0}^{M} a_j(\Theta)\mathbb{T}_j(2t-1), \tag{5.10}$$

$$z_M(t;\Theta) = (1-t) \sum_{j=0}^{M} b_j(\Theta)\mathbb{T}_j(2t-1), \tag{5.11}$$

$$w_M(t;\Theta) = \sum_{j=0}^{M} c_j(\Theta)\mathbb{T}_j(2t-1). \tag{5.12}$$

Hence, the prescribed conditions $q_M(0;\Theta) = 0$ and $z_M(1;\Theta) = 0$ are satisfied exactly for every admissible parameter set Θ, and no separate boundary or transversality penalty is required in this example. For each spectral degree M, the PI-CSNN is trained independently by minimizing the weighted residual loss, while the analytical expressions in (5.3) are used only after training to compute the L^2 and L^{∞} errors. We run simulations for the spectral degrees $M = 4, \ldots, 20$ for a comparison of the convergence of the proposed PI-CSNN framework for $\alpha = 16/3$, $8/3$, and $2/3$ to investigate the impact of the regularity of the solution. For each M, the trainable neural architecture produces the global Chebyshev coefficient vectors that define the state and control approximations. The error curves are presented in Figures 2 and 3 while the typical L^2 and L^∞ errors are listed in Tables 1 and 2. For the smooth case $\alpha = 16/3$, we observe a fast decay of the error. As M increases from 4 to 20, the L^2 errors decrease overall from

$$2.266 \times 10^{-4} \text{ to } 3.063 \times 10^{-12}$$

for the state, and

$$1.885 \times 10^{-3} \text{ to } 2.426 \times 10^{-12}$$

for the control. The corresponding L^∞ errors decrease overall from

$$5.612 \times 10^{-4} \text{ to } 1.591 \times 10^{-11} \quad \text{and} \quad 6.234 \times 10^{-3} \text{ to } 3.277 \times 10^{-11},$$

respectively. These results demonstrate the high-order approximation capability of the PI-CSNN framework for sufficiently smooth solutions. For $\alpha = 8/3$, the convergence remains systematic overall but is slower because of reduced regularity. The L^2 state and control errors decrease overall from

$$4.111 \times 10^{-5} \text{ and } 1.061 \times 10^{-4} \text{ at } M = 4$$

to

$$1.493 \times 10^{-8} \text{ and } 1.735 \times 10^{-8} \text{ at } M = 20,$$

respectively, while the L^∞ errors decrease from

$$8.999 \times 10^{-5} \text{ and } 4.955 \times 10^{-4}$$

to

$$6.948 \times 10^{-8} \text{ and } 3.046 \times 10^{-7},$$

respectively, confirming that high accuracy is retained under moderately reduced regularity. For the weakly regular case $\alpha = 2/3$, limited endpoint regularity leads to slower convergence. Nevertheless, the L^2 errors decrease overall from

$$5.929 \times 10^{-3} \text{ to } 3.790 \times 10^{-4}$$

for the state, and from

$$4.764 \times 10^{-3} \text{ to } 8.406 \times 10^{-4}$$

for the control. The corresponding L^∞ errors are reduced overall from

$$2.261 \times 10^{-2} \text{ to } 3.697 \times 10^{-3}$$

and from

$$3.616 \times 10^{-2} \text{ to } 1.682 \times 10^{-2},$$

respectively. Minor non-monotone variations at some spectral degrees are observed, particularly in the control error, and are consistent with the combined influence of limited endpoint regularity, finite-dimensional spectral approximation, and neural optimization. The results clearly show a general regularity-dependent convergence pattern. The fastest decay of error is for $\alpha = 16/3$ and then $\alpha = 8/3$ while the weakly regular case $\alpha = 2/3$ converges slower. For the smooth benchmark, the L^2 errors are on the order of 10^{-12} and the L^∞ errors are on the order of 10^{-11}. Also the overall error reduction is substantially preserved for the lower regularity case. The results show that the PI-CSNN framework can successfully combine the physics-informed neural generation of global spectral coefficients with the high-order approximation ability of shifted Chebyshev expansions.

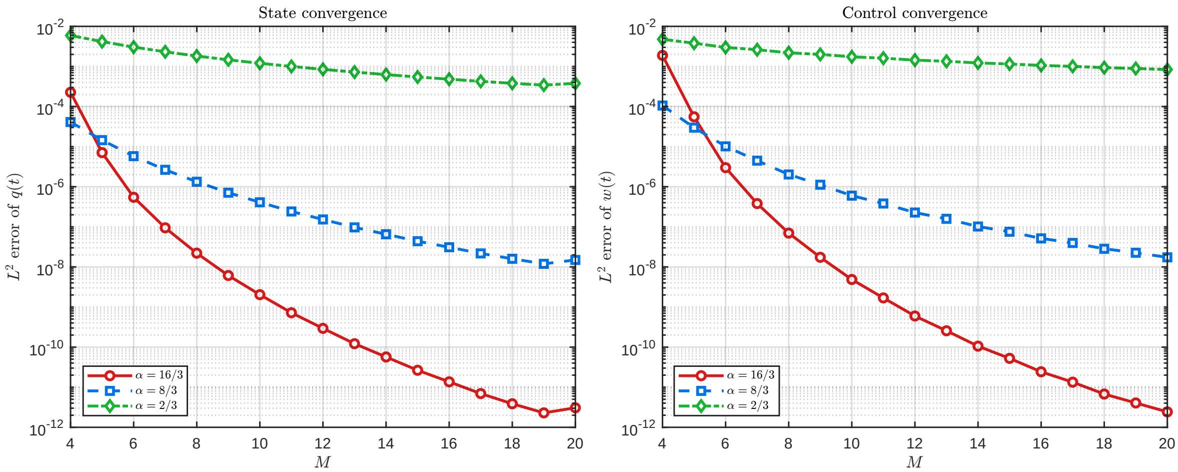

Figure 2. Convergence of the proposed PI-CSNN approximation in the L^2 norm for Example 1 with different values of α.

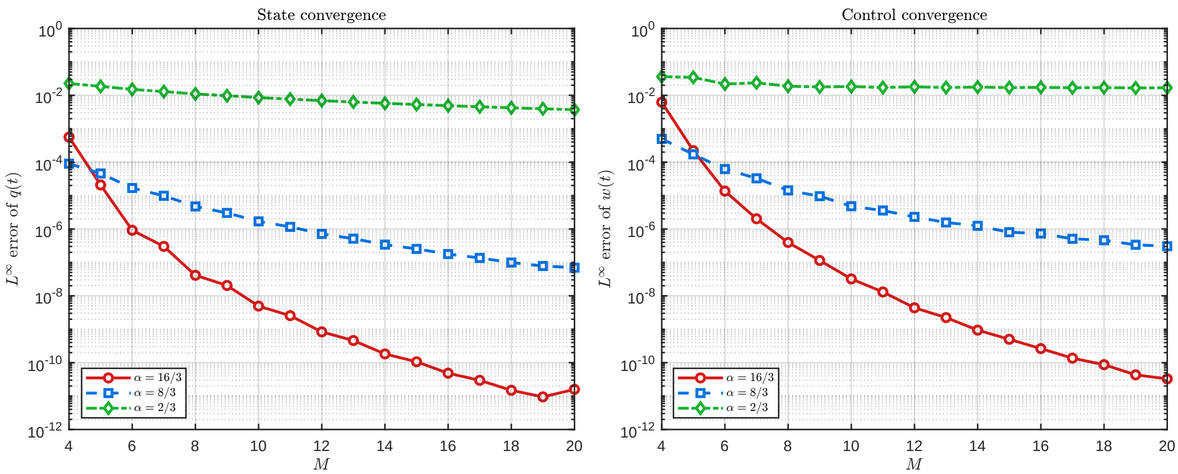

Figure 3. Convergence of the proposed PI-CSNN approximation in the L^∞ norm for Example 1 with different values of α.

Table 1. L^2 errors of the proposed PI-CSNN approximation for the state and control variables in Example 1.

α	M	L^2 error (state)	L^2 error (control)
16/3	4	2.266×10^{-4}	1.885×10^{-3}
16/3	8	2.204×10^{-8}	6.985×10^{-8}
16/3	12	2.932×10^{-10}	5.968×10^{-10}
16/3	16	1.363×10^{-11}	2.426×10^{-11}
16/3	20	3.063×10^{-12}	2.426×10^{-12}
8/3	4	4.111×10^{-5}	1.061×10^{-4}
8/3	8	1.333×10^{-6}	2.025×10^{-6}
8/3	12	1.526×10^{-7}	2.273×10^{-7}
8/3	16	3.096×10^{-8}	5.146×10^{-8}
8/3	20	1.493×10^{-8}	1.735×10^{-8}
2/3	4	5.929×10^{-3}	4.764×10^{-3}
2/3	8	1.820×10^{-3}	2.197×10^{-3}
2/3	12	8.473×10^{-4}	1.439×10^{-3}
2/3	16	4.809×10^{-4}	1.061×10^{-3}
2/3	20	3.790×10^{-4}	8.406×10^{-4}

Table 2. L^∞ errors of the proposed PI-CSNN approximation for the state and control variables in Example 1.

α	M	L^∞ error (state)	L^∞ error (control)
16/3	4	5.612×10^{-4}	6.234×10^{-3}
16/3	8	4.120×10^{-8}	3.951×10^{-7}
16/3	12	8.343×10^{-10}	4.377×10^{-9}
16/3	16	4.844×10^{-11}	2.638×10^{-10}
16/3	20	1.591×10^{-11}	3.277×10^{-11}
8/3	4	8.999×10^{-5}	4.955×10^{-4}
8/3	8	4.716×10^{-6}	1.429×10^{-5}
8/3	12	7.138×10^{-7}	2.302×10^{-6}
8/3	16	1.775×10^{-7}	7.339×10^{-7}
8/3	20	6.948×10^{-8}	3.046×10^{-7}
2/3	4	2.261×10^{-2}	3.616×10^{-2}
2/3	8	1.101×10^{-2}	1.872×10^{-2}
2/3	12	6.908×10^{-3}	1.793×10^{-2}
2/3	16	4.888×10^{-3}	1.729×10^{-2}
2/3	20	3.697×10^{-3}	1.682×10^{-2}

Example 2. We next consider a distributed-order fractional time-invariant optimal control problem to evaluate the performance of the proposed PI-CSNN framework for computing the state, adjoint, and control variables. This example assesses the accuracy of the computed optimal cost and the satisfaction of the distributed-order optimality system under different fractional-order distributions. The problem

is defined as follows:

$$\min_{w} J(q,w) = \frac{1}{2}\int_0^1 \left(q^2(t) + w^2(t)\right) dt, \tag{5.13}$$

subject to

$${}^{C}\mathfrak{D}_0^{(p)} q(t) = w(t) - q(t), \quad q(0) = 1, \tag{5.14}$$

where $p(\mu)$ denotes the fractional-order distribution on $[0, 1]$.

The Hamiltonian is

$$\mathcal{H}(q,w,z) = \frac{1}{2}\left(q^2 + w^2\right) + z(t)\left(w(t) - q(t)\right). \tag{5.15}$$

The stationarity condition yields $w(t) = -z(t)$, leading to the reduced state-adjoint system

$${}^{C}\mathfrak{D}_0^{(p)} q(t) + q(t) + z(t) = 0, \qquad q(0) = 1, \tag{5.16}$$

$${}^{RL}\mathfrak{D}_1^{(p)} z(t) - q(t) + z(t) = 0, \qquad z(1) = 0. \tag{5.17}$$

The control is recovered from $w = -z$; hence, no independent control output or separate optimality residual is required. The following configurations are considered:
Case I:

$$p(\mu) = \delta(\mu - 1), \tag{5.18}$$

Case II:

$$p(\mu) = \delta(\mu - 0.99), \tag{5.19}$$

Case III:

$$p(\mu) = \delta(\mu - 0.9), \tag{5.20}$$

Case IV:

$$p(\mu) = \delta(\mu - 0.8), \tag{5.21}$$

Case V:

$$p(\mu) = N(0.8, 0.05)(\mu). \tag{5.22}$$

Cases I–IV are evaluated directly at the prescribed fractional orders. For Case V, the Gaussian density is restricted to $[0, 1]$, normalized to unit mass, and integrated using $K = 40$ Gauss-Legendre order-quadrature nodes. For Case I, the operators reduce to their classical first-order counterparts, and the analytical state and control are

$$q(t) = C\sinh(\sqrt{2}\,t) + \cosh(\sqrt{2}\,t), \tag{5.23}$$

$$w(t) = (C + \sqrt{2})\sinh(\sqrt{2}\,t) + (C\sqrt{2} + 1)\cosh(\sqrt{2}\,t), \tag{5.24}$$

where

$$C = -\frac{2\sinh(\sqrt{2}) + \sqrt{2}\cosh(\sqrt{2})}{2\cosh(\sqrt{2}) + \sqrt{2}\sinh(\sqrt{2})}. \tag{5.25}$$

The corresponding adjoint is $z = -w$. These expressions are used only for post-training validation. The boundary-adapted PI-CSNN approximations are

$$q_M(t;\Theta) = 1 + t\sum_{j=0}^{M} a_j(\Theta)\mathbb{T}_j(2t-1), \tag{5.26}$$

$$z_M(t;\Theta) = (1-t)\sum_{j=0}^{M} b_j(\Theta)\mathbb{T}_j(2t-1), \tag{5.27}$$

with

$$w_M(t;\Theta) = -z_M(t;\Theta), \tag{5.28}$$

so that $q_M(0;\Theta) = 1$ and $z_M(1;\Theta) = 0$ hold exactly. The corresponding residuals are

$$\mathcal{R}_q(t_i;\Theta) = {}^{C}\mathfrak{D}_0^{(p)}\, q_M(t_i;\Theta) + q_M(t_i;\Theta) + z_M(t_i;\Theta), \tag{5.29}$$

$$\mathcal{R}_z(t_i;\Theta) = {}^{RL}\mathfrak{D}_1^{(p)}\, z_M(t_i;\Theta) - q_M(t_i;\Theta) + z_M(t_i;\Theta). \tag{5.30}$$

The network parameters are obtained by minimizing the associated weighted residual loss. The numerical state, adjoint, and control profiles for the five cases are displayed in Figure 4, while Figure 5 illustrates the convergence behavior with increasing spectral degree M. The profiles remain smooth across the single-order and genuinely distributed-order configurations and vary systematically with the prescribed-order distribution. For Case I, the analytical solution provides an independent validation of the trained PI-CSNN approximation. At $M = 15$, the post-training errors obtained from the validation run are

$$\begin{aligned} \|q - q_M\|_{L^2(0,1)} &= 1.612\times 10^{-15}, & \|q - q_M\|_{L^\infty(0,1)} &= 4.885\times 10^{-15},\\ \|z - z_M\|_{L^2(0,1)} &= 1.244\times 10^{-15}, & \|z - z_M\|_{L^\infty(0,1)} &= 3.886\times 10^{-15},\\ \|w - w_M\|_{L^2(0,1)} &= 1.244\times 10^{-15}, & \|w - w_M\|_{L^\infty(0,1)} &= 3.886\times 10^{-15}. \end{aligned} \tag{5.31}$$

The corresponding optimal cost is

$$J_{\mathrm{I}} = 0.192909298093, \tag{5.32}$$

with an absolute difference of 5.551×10^{-17} from the analytical value. The trained solution also gives

$$\|\mathcal{R}_q\|_{L^2} = 4.578\times 10^{-14}, \qquad \|\mathcal{R}_z\|_{L^2} = 3.792\times 10^{-14}, \tag{5.33}$$

together with

$$\|\mathcal{R}_q\|_{L^\infty} = 2.024\times 10^{-13}, \qquad \|\mathcal{R}_z\|_{L^\infty} = 2.635\times 10^{-13}. \tag{5.34}$$

These values confirm highly accurate satisfaction of the governing state-adjoint system. The computed cost is consistent with the benchmark values reported for the variational iteration method (VIM), Adomian decomposition method (ADM), Legendre wavelet method (LWM), Chebyshev wavelet method (CWM), Laguerre wavelet method (LaWM), and cosine and sine wavelet method (CASWM) [37–39], as well as the Legendre collocation method (LCM) [15], as summarized in Table 3.

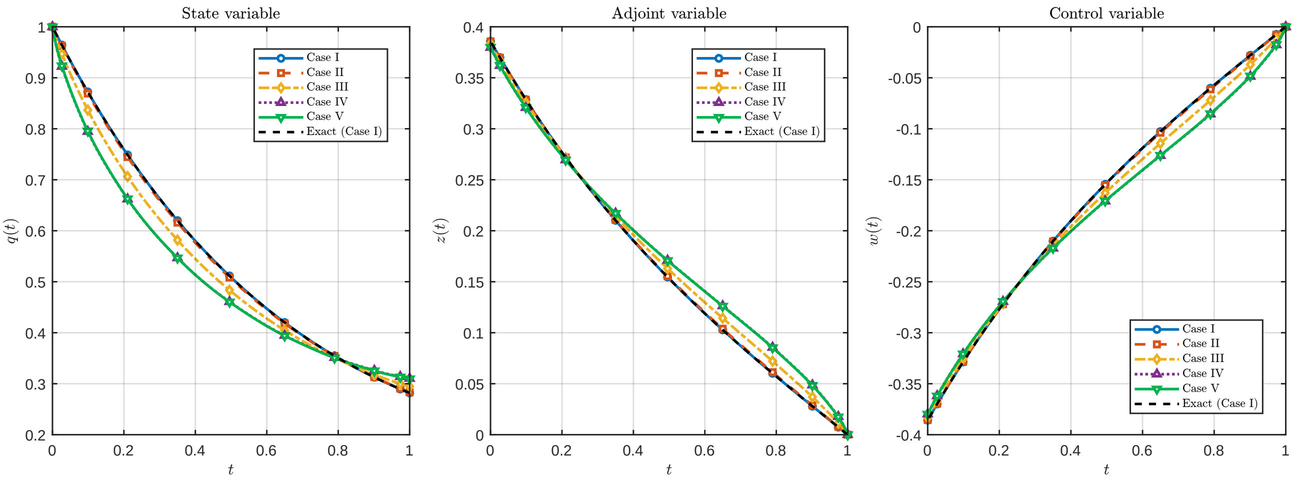

Figure 4. PI-CSNN state, adjoint, and control profiles for the five cases of Example 2.

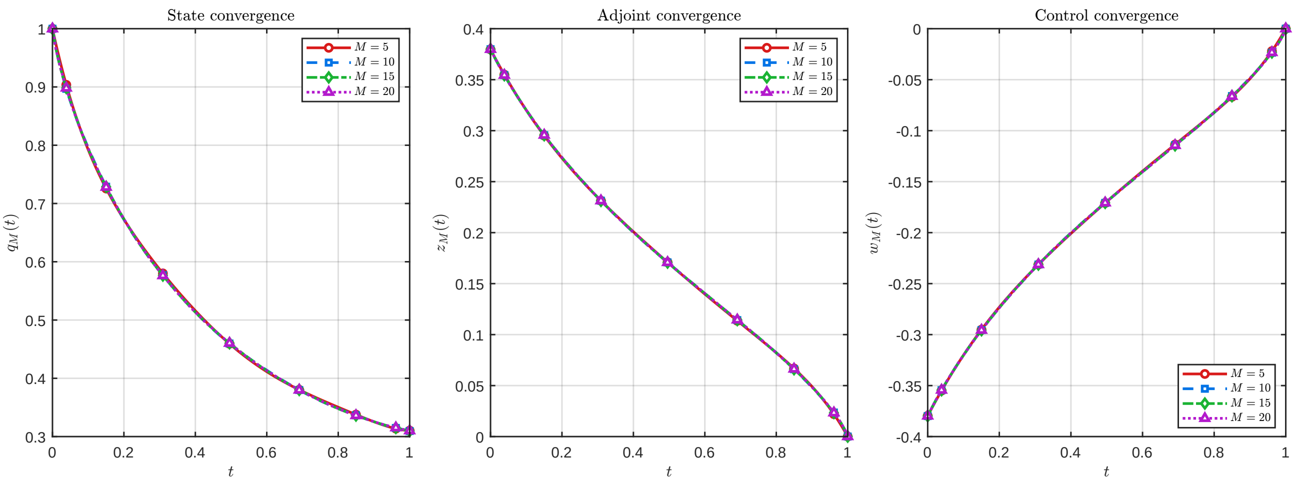

Figure 5. PI-CSNN convergence for the distributed-order Case V of Example 2 with increasing spectral degree M.

Table 3. Comparison of optimal costs for the distributed-order fractional time-invariant optimal control problem under different fractional-order distributions with $M = 15$ in Example 2.

$p(\mu)$	VIM	ADM	LWM	CWM	LaWM	CASWM	LCM	Proposed PI-CSNN
Case I	0.192909	0.192909	0.192909	0.192921	0.192604	0.196914	0.192909	0.192909298093
Case II	–	–	–	–	–	–	–	0.191530902205
Case III	0.179530	0.179620	0.179529	0.179641	0.185942	0.180555	0.179690	0.179546635955
Case IV	0.167110	0.167400	0.167078	0.167328	0.181798	0.168072	0.167347	0.167147884059
Case V	–	–	–	–	–	–	0.165906	0.167188260264

The separate Case I spectral-degree study exhibits rapid error decay as M increases. The L^2 state error decreases from

$$2.691 \times 10^{-6} \text{ at } M = 4 \text{ to } 8.004 \times 10^{-12} \text{ at } M = 8,$$

while the corresponding control error decreases from

$$8.116 \times 10^{-7} \text{ to } 2.411 \times 10^{-12}.$$

At $M = 10$, the state and control errors reach

$$7.545 \times 10^{-15} \text{ and } 2.273 \times 10^{-15},$$

respectively. At $M = 12$, they are further reduced to

$$3.282 \times 10^{-16} \text{ and } 3.045 \times 10^{-16},$$

respectively. Beyond this point, the errors remain approximately at the floating-point accuracy level, with small non-monotone variations consistent with numerical roundoff and neural optimization.

The L^∞ errors exhibit the same rapid decay. In particular, the state and control errors decrease from

$$4.241 \times 10^{-6} \text{ and } 1.279 \times 10^{-6} \text{ at } M = 4$$

to

$$1.271 \times 10^{-11} \text{ and } 3.833 \times 10^{-12} \text{ at } M = 8,$$

respectively. This behavior demonstrates rapid high-order convergence for the smooth classical benchmark. The cost approximation in Case I also converges rapidly. The absolute cost difference decreases from

$$2.539 \times 10^{-9} \text{ at } M = 4 \text{ to } 9.984 \times 10^{-12} \text{ at } M = 5$$

and

$$3.003 \times 10^{-14} \text{ at } M = 6.$$

For larger spectral degrees, the cost difference reaches values near machine precision, including zero for the reported numerical precision at $M = 10$. For Cases II–IV, corresponding to the single fractional orders 0.99, 0.9, and 0.8, respectively, the PI-CSNN yields the optimal costs

$$J_{\text{II}} = 0.191530902205, \qquad J_{\text{III}} = 0.179546635955, \qquad J_{\text{IV}} = 0.167147884059. \tag{5.35}$$

The corresponding L^2 residual norms are

$$\begin{aligned} \text{Case II:} \quad & \|\mathcal{R}_q\|_{L^2} = 8.470 \times 10^{-4}, \quad \|\mathcal{R}_z\|_{L^2} = 1.746 \times 10^{-4}, \\ \text{Case III:} \quad & \|\mathcal{R}_q\|_{L^2} = 8.551 \times 10^{-3}, \quad \|\mathcal{R}_z\|_{L^2} = 1.847 \times 10^{-3}, \\ \text{Case IV:} \quad & \|\mathcal{R}_q\|_{L^2} = 1.717 \times 10^{-2}, \quad \|\mathcal{R}_z\|_{L^2} = 3.938 \times 10^{-3}. \end{aligned} \tag{5.36}$$

The progressive decrease in the optimal cost reflects the influence of the fractional order on the nonlocal temporal dynamics. These values follow the established benchmark trend and are compared with the available numerical methods in Table 3. Case V represents the genuinely distributed-order configuration with a normalized Gaussian density centered at $\mu = 0.8$ and standard deviation 0.05. At $M = 15$, the trained PI-CSNN approximation gives $J_{\text{V}} = 0.167188260264$. The corresponding residual norms are

$$\|\mathcal{R}_q\|_{L^2} = 1.579 \times 10^{-2}, \qquad \|\mathcal{R}_z\|_{L^2} = 3.614 \times 10^{-3} \tag{5.37}$$

and

$$\|\mathcal{R}_q\|_{L^\infty} = 6.923 \times 10^{-1}, \qquad \|\mathcal{R}_z\|_{L^\infty} = 1.553 \times 10^{-1}. \tag{5.38}$$

The spectral-degree study for Case V provides an additional assessment of the genuinely distributed-order approximation. For $M = 5, 10, 15$, and 20, the computed costs are

$$\begin{aligned} J_{\mathrm{V}}^{(5)} &= 0.167633740767, & J_{\mathrm{V}}^{(10)} &= 0.167260972425, \\ J_{\mathrm{V}}^{(15)} &= 0.167188260265, & J_{\mathrm{V}}^{(20)} &= 0.167174767630. \end{aligned} \tag{5.39}$$

Over the same sequence of spectral degrees, the state residual norms are

$$\begin{aligned} \|\mathcal{R}_q^{(5)}\|_{L^2} &= 4.494 \times 10^{-2}, & \|\mathcal{R}_q^{(10)}\|_{L^2} &= 2.349 \times 10^{-2}, \\ \|\mathcal{R}_q^{(15)}\|_{L^2} &= 1.579 \times 10^{-2}, & \|\mathcal{R}_q^{(20)}\|_{L^2} &= 1.188 \times 10^{-2}, \end{aligned} \tag{5.40}$$

while the corresponding adjoint residual norms are

$$\begin{aligned} \|\mathcal{R}_z^{(5)}\|_{L^2} &= 1.130 \times 10^{-2}, & \|\mathcal{R}_z^{(10)}\|_{L^2} &= 5.102 \times 10^{-3}, \\ \|\mathcal{R}_z^{(15)}\|_{L^2} &= 3.614 \times 10^{-3}, & \|\mathcal{R}_z^{(20)}\|_{L^2} &= 2.639 \times 10^{-3}. \end{aligned} \tag{5.41}$$

Thus, both residual norms decrease systematically as the spectral dimension increases, while the changes in the computed cost become progressively smaller. This behavior indicates improved satisfaction of the distributed-order state-adjoint system together with stabilization of the objective value. Overall, the computed optimal costs for Cases I–V are

$$\begin{gathered} J_{\mathrm{I}} = 0.192909298093, \quad J_{\mathrm{II}} = 0.191530902205, \quad J_{\mathrm{III}} = 0.179546635955, \\ J_{\mathrm{IV}} = 0.167147884059, \quad J_{\mathrm{V}} = 0.167188260264. \end{gathered} \tag{5.42}$$

The near-machine-precision recovery of the classical Case I analytical benchmark, the agreement of the single-order fractional cases with the established benchmark trends, and the systematic residual reduction in the genuinely distributed-order Case V provide complementary numerical evidence for the accuracy and applicability of the PI-CSNN framework to time-invariant DOFOCPs. The computed state, adjoint, and control profiles for the five fractional-order configurations are displayed in Figure 4. The corresponding L^2 and L^∞ errors for the analytical Case I benchmark at $M = 15$ are summarized in Table 4. The residual norms of the trained PI-CSNN solutions for the five fractional-order configurations at $M = 15$ are summarized in Table 5. The spectral-degree convergence of the genuinely distributed-order Case V is illustrated in Figure 5, while the rapid convergence of the Case I solution with increasing spectral degree M is illustrated in Figure 6.

Table 4. L^2 and L^∞ errors for the analytical Case I benchmark of Example 2 obtained using the proposed PI-CSNN framework at $M = 15$.

Variable	L^2 error	L^∞ error
State variable	1.612×10^{-15}	4.885×10^{-15}
Adjoint variable	1.244×10^{-15}	3.886×10^{-15}
Control variable	1.244×10^{-15}	3.886×10^{-15}

Table 5. Residual norms of the trained PI-CSNN solutions for the five fractional-order configurations in Example 2 at $M = 15$.

Case	$\|\mathcal{R}_q\|_{L^2}$	$\|\mathcal{R}_z\|_{L^2}$	$\|\mathcal{R}_q\|_{L^\infty}$	$\|\mathcal{R}_z\|_{L^\infty}$
Case I	4.578×10^{-14}	3.792×10^{-14}	2.024×10^{-13}	2.635×10^{-13}
Case II	8.470×10^{-4}	1.746×10^{-4}	4.605×10^{-2}	9.398×10^{-3}
Case III	8.551×10^{-3}	1.847×10^{-3}	4.219×10^{-1}	8.986×10^{-2}
Case IV	1.717×10^{-2}	3.938×10^{-3}	7.464×10^{-1}	1.678×10^{-1}
Case V	1.579×10^{-2}	3.614×10^{-3}	6.923×10^{-1}	1.553×10^{-1}

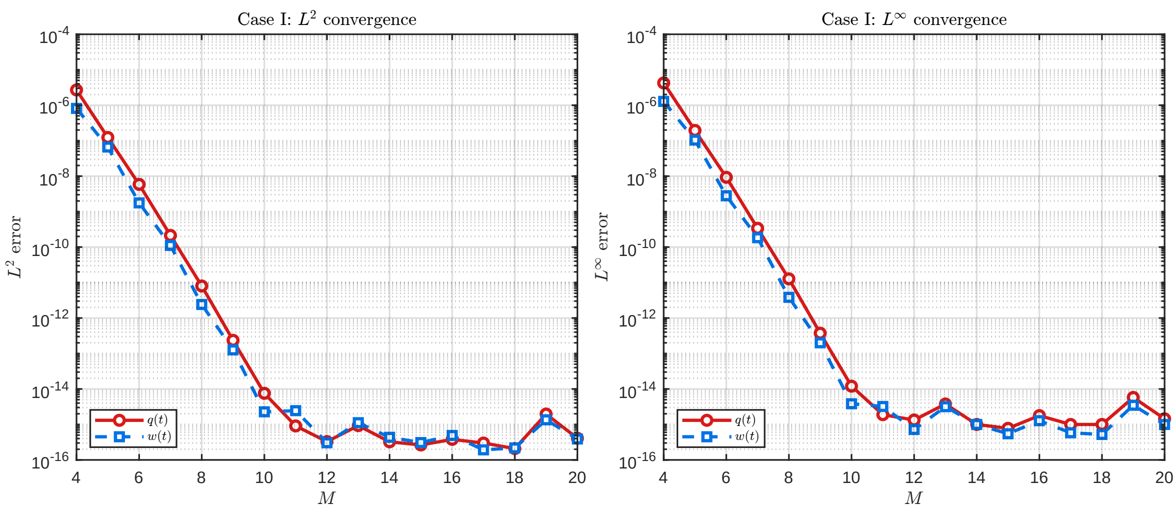

Figure 6. PI-CSNN convergence for Case I of Example 2 with increasing spectral degree M.

To verify that the numerical accuracy is not limited by the approximation of the distributed-order integral, a quadrature sensitivity study was performed for Case V, which involves the continuous normal density $\mathcal{N}(0.8, 0.05)$. Throughout this experiment, the Chebyshev spectral degree, the temporal collocation points, the PI-CSNN architecture, and all optimization parameters were fixed at $M = 15$, $N_c = 180$, and the values used in the previous experiments. Only the number K of Gauss-Legendre quadrature nodes employed for the distributed-order integration was varied. The solution obtained with $K = 60$ was taken as the high-resolution reference.

Table 6 demonstrates that the numerical solution is sensitive to the quadrature resolution only for relatively small values of K. As K increases from 5 to 17, the relative change in the optimal cost decreases from 2.12×10^{-2} to 8.73×10^{-5}, while the L^2 differences in the state and adjoint variables decrease by several orders of magnitude. For $K \geq 25$, both the relative change in the optimal cost and the solution differences are of the order of 10^{-8} or smaller, indicating that the distributed-order quadrature has effectively converged. Moreover, the physics-informed residual norms remain nearly unchanged for $K \geq 25$, confirming that the remaining numerical error is dominated by the spectral approximation and optimization processes rather than by the distributed-order quadrature. Therefore, the choice $K = 40$ adopted in all numerical experiments provides a reliable and conservative balance between computational efficiency and quadrature accuracy. The reference solution J_{ref} is computed using $K_{\text{ref}} = 60$.

Table 6. Quadrature sensitivity results for Case V of Example 2.

K	Optimal cost J_K	$\dfrac{\lvert J_K - J_{\text{ref}}\rvert}{\lvert J_{\text{ref}}\rvert}$	$\lVert q_K - q_{\text{ref}}\rVert_{L^2}$	$\lVert z_K - z_{\text{ref}}\rVert_{L^2}$	$\lVert\mathcal{R}_q\rVert_{L^2}$	$\lVert\mathcal{R}_z\rVert_{L^2}$
5	0.163649193188	2.117×10^{-2}	8.378×10^{-3}	2.956×10^{-3}	1.955×10^{-2}	4.570×10^{-3}
9	0.168071717592	5.284×10^{-3}	1.897×10^{-3}	5.459×10^{-4}	1.593×10^{-2}	3.634×10^{-3}
13	0.167357293536	1.011×10^{-3}	3.990×10^{-4}	1.407×10^{-4}	1.565×10^{-2}	3.579×10^{-3}
17	0.167173652593	8.733×10^{-5}	3.223×10^{-5}	1.011×10^{-5}	1.579×10^{-2}	3.615×10^{-3}
25	0.167188258166	3.037×10^{-8}	2.315×10^{-8}	4.774×10^{-9}	1.579×10^{-2}	3.614×10^{-3}
40	0.167188260265	4.293×10^{-8}	5.186×10^{-8}	1.560×10^{-8}	1.579×10^{-2}	3.614×10^{-3}
60	0.167188253087	0	0	0	1.579×10^{-2}	3.614×10^{-3}

Example 3. Finally, we consider a distributed-order fractional time-varying optimal control problem to further evaluate the robustness and applicability of the proposed PI-CSNN framework. This example examines the ability of the proposed method to accurately approximate the state, adjoint, and control variables and to compute the corresponding optimal cost for a time-dependent distributed-order system under different fractional-order distributions. The problem is formulated as follows:

$$\min_{w} J(q,w) = \frac{1}{2}\int_0^1 (q^2(t) + w^2(t))\,dt, \tag{5.43}$$

subject to

$${}^{C}\mathfrak{D}_0^{(p)}\, q(t) = w(t) + tq(t), \quad q(0) = 1. \tag{5.44}$$

Compared with Example 2, the term $tq(t)$ introduces explicit time dependence into the state dynamics. The Hamiltonian is

$$\mathcal{H}(q,w,z) = \frac{1}{2}(q^2 + w^2) + z(t)\,(w(t) + tq(t)) \tag{5.45}$$

and the stationarity condition again gives $w(t) = -z(t)$. The reduced state-adjoint system is therefore

$${}^{C}\mathfrak{D}_0^{(p)}\, q(t) + z(t) - tq(t) = 0, \quad q(0) = 1, \tag{5.46}$$

$${}^{RL}\mathfrak{D}_1^{(p)}\, z(t) - tz(t) - q(t) = 0, \quad z(1) = 0. \tag{5.47}$$

As in Example 2, the control is recovered from $w = -z$. The five fractional-order configurations defined in (5.22) are retained, with Cases I–IV evaluated directly at orders 1, 0.99, 0.9, and 0.8, respectively, and Case V treated as the normalized Gaussian distributed-order configuration. Using the same reduced-control representation as in Example 2, the PI-CSNN approximations are

$$q_M(t;\Theta) = 1 + t\sum_{j=0}^{M} a_j(\Theta)\mathbb{T}_j(2t-1), \tag{5.48}$$

$$z_M(t;\Theta) = (1-t)\sum_{j=0}^{M} b_j(\Theta)\mathbb{T}_j(2t-1), \tag{5.49}$$

with

$$w_M(t;\Theta) = -z_M(t;\Theta). \tag{5.50}$$

These forms enforce $q_M(0;\Theta) = 1$ and $z_M(1;\Theta) = 0$ exactly. The problem-specific residuals are

$$\mathcal{R}_q(t_i;\Theta) = {}^{C}\mathfrak{D}_0^{(p)}\, q_M(t_i;\Theta) + z_M(t_i;\Theta) - t_i q_M(t_i;\Theta), \tag{5.51}$$

$$\mathcal{R}_z(t_i;\Theta) = {}^{RL}\mathfrak{D}_1^{(p)}\, z_M(t_i;\Theta) - t_i z_M(t_i;\Theta) - q_M(t_i;\Theta). \tag{5.52}$$

The trainable parameters are determined from the weighted residual loss using the optimization strategy specified in the general PI-CSNN methodology. The numerical state, adjoint, and control profiles for the five fractional-order configurations are displayed in Figure 7. The computed solutions remain smooth and vary systematically with the prescribed-order distribution, illustrating the influence of fractional memory on the time-varying optimal dynamics. The corresponding optimal costs and residual norms at $M = 15$ are summarized in Table 7.

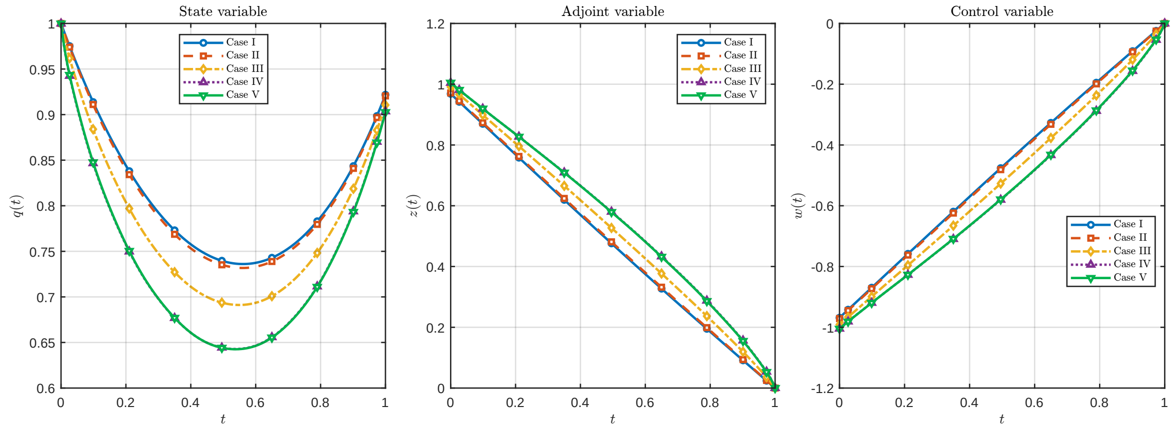

Figure 7. PI-CSNN state, adjoint, and control profiles for the five cases of Example 3.

Table 7. Optimal costs and residual norms obtained by the proposed PI-CSNN framework for Example 3 at $M = 15$.

Case	Optimal cost	$\|\mathcal{R}_q\|_{L^2}$	$\|\mathcal{R}_z\|_{L^2}$
I	0.484267696229	1.571×10^{-15}	2.748×10^{-15}
II	0.483462914405	5.921×10^{-4}	5.599×10^{-4}
III	0.475892125919	6.087×10^{-3}	5.574×10^{-3}
IV	0.467008616133	1.234×10^{-2}	1.104×10^{-2}
V	0.466726603779	1.135×10^{-2}	1.015×10^{-2}

For Case I, the PI-CSNN yields an optimal cost of 0.484267696229, with state and adjoint L^2 residual norms of 1.571×10^{-15} and 2.748×10^{-15}, respectively. The corresponding L^∞ residual norms are 1.821×10^{-14} and 1.910×10^{-14}. These values are close to machine precision and provide a stringent verification of the state-adjoint implementation for the classical first-order configuration. For Cases II–IV, the computed optimal costs are

$$J_{\mathrm{II}} = 0.483462914405, \quad J_{\mathrm{III}} = 0.475892125919, \quad J_{\mathrm{IV}} = 0.467008616133. \tag{5.53}$$

The corresponding L^2 residual pairs $(\|\mathcal{R}_q\|_{L^2}, \|\mathcal{R}_z\|_{L^2})$ are

$$\left(5.921 \times 10^{-4}, 5.599 \times 10^{-4}\right), \quad \left(6.087 \times 10^{-3}, 5.574 \times 10^{-3}\right), \quad \left(1.234 \times 10^{-2}, 1.104 \times 10^{-2}\right), \tag{5.54}$$

respectively. The systematic variation in the computed costs and solution profiles reflects the influence of the fractional order on the time-varying optimal dynamics. For the genuinely distributed-order Case V, the proposed PI-CSNN framework gives $J_{\mathrm{V}} = 0.466726603779$, with

$$\|\mathcal{R}_q\|_{L^2} = 1.135 \times 10^{-2}, \qquad \|\mathcal{R}_z\|_{L^2} = 1.015 \times 10^{-2}. \tag{5.55}$$

The corresponding L^∞ residual norms are 5.020×10^{-1} and 4.487×10^{-1}, respectively. To examine the effect of the spectral degree on the genuinely distributed-order configuration, independent PI-CSNN computations are performed for $M = 5$, 10, 15, and 20. The results are reported in Table 8 and illustrated in Figure 8.

Table 8. Spectral-degree study of the PI-CSNN framework for the genuinely distributed-order Case V of Example 3.

M	Optimal cost	$\|\mathcal{R}_q\|_{L^2}$	$\|\mathcal{R}_z\|_{L^2}$
5	0.466937537857	3.090×10^{-2}	2.738×10^{-2}
10	0.466759355488	1.689×10^{-2}	1.459×10^{-2}
15	0.466726603778	1.135×10^{-2}	1.015×10^{-2}
20	0.466744588275	8.602×10^{-3}	7.626×10^{-3}

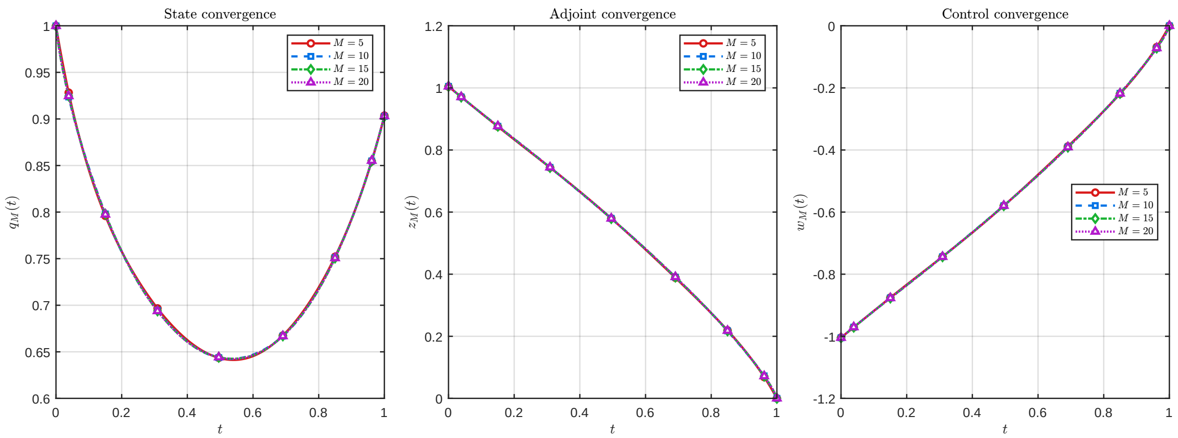

Figure 8. PI-CSNN convergence profiles for the distributed-order Case V of Example 3 with increasing spectral degree M.

As M increases from 5 to 20, the state residual decreases systematically from

$$3.090 \times 10^{-2} \text{ to } 8.602 \times 10^{-3},$$

while the adjoint residual decreases from

$$2.738 \times 10^{-2} \text{ to } 7.626 \times 10^{-3}.$$

The corresponding optimal costs for $M = 5$, 10, 15, and 20 are

$$0.466937537857, \; 0.466759355488, \; 0.466726603778, \; \text{and } 0.466744588275,$$

respectively. Thus, although the cost is not strictly monotone with respect to M, both residual norms decrease systematically, while the objective value stabilizes near 0.4667. This behavior indicates improved satisfaction of the governing state-adjoint system as the spectral dimension increases, together with stabilization of the computed objective value. The optimal costs at $M = 15$ are compared in Table 9 with available results from the shifted Jacobi method (SJM) [40], LSM, VIM, LWM, CWM, LaWM, and CASWM. The near-machine-precision residuals obtained for Case I provide a stringent implementation check, while the fractional cases follow the benchmark trends of the comparison methods. For Case V, although the computed cost differs from the reported LSM value, the independent spectral-degree study shows systematic reduction of both residual norms and stabilization of the objective value.

Table 9. Comparison of optimal costs for the distributed-order fractional time-varying optimal control problem under different fractional-order distributions with $M = 15$ in Example 3.

$p(\mu)$	SJM	LSM	VIM	LWM	CWM	LaWM	CASWM	Proposed PI-CSNN
Case I	0.484260	0.484268	0.484260	0.484268	0.484232	0.483476	0.500511	0.484267696229
Case II	–	–	–	–	–	–	–	0.483462914405
Case III	0.475880	0.476446	0.475930	0.475886	0.475881	0.473351	0.492513	0.475892125919
Case IV	0.466970	0.460500	0.467220	0.466993	0.467005	0.459714	0.483723	0.467008616133
Case V	–	0.451530	–	–	–	–	–	0.466726603779

Overall, the results demonstrate that the proposed PI-CSNN framework captures the influence of the fractional-order distribution on the time-varying optimal dynamics. The near-machine-precision residuals for the classical configuration, the systematic variation across the single-order fractional cases, and the decreasing residual norms for the genuinely distributed-order case provide complementary numerical evidence for the accuracy and consistency of the proposed implementation.

To examine the influence of the distributed-order quadrature resolution on the numerical solution of Example 3, a quadrature sensitivity study was performed for Case V, corresponding to the continuous normal density $\mathcal{N}(0.8, 0.05)$. The spectral degree, temporal collocation resolution, PI-CSNN architecture, and optimization parameters were fixed at $M = 15$ and $N_t = 180$, while only the number K of Gauss-Legendre quadrature nodes was varied. The solution obtained with $K_{\text{ref}} = 60$ was used as the high-resolution reference. The reported quantities include the optimal cost, the relative change in the cost, the L^2 differences in the state and adjoint variables, and the state and adjoint residual norms.

As shown in Table 10, the quadrature approximation has a noticeable influence only for relatively small values of K. When K increases from 5 to 17, the relative change in the optimal cost decreases from 5.191×10^{-3} to 2.726×10^{-5}. Over the same range, the L^2 differences in the state and adjoint variables decrease from approximately 10^{-2} to 10^{-5}. For $K \geq 25$, the relative cost variation and the solution differences are of order 10^{-8}, indicating that the distributed-order Gauss-Legendre quadrature has effectively converged. For $K \geq 25$, the relative variation in the optimal cost is below 4×10^{-8}, while the state and adjoint differences are also of order 10^{-8}. In contrast, the corresponding physics-informed residual norms remain approximately

$$\|\mathcal{R}_q\|_{L^2} \approx 1.135 \times 10^{-2} \quad \text{and} \quad \|\mathcal{R}_z\|_{L^2} \approx 1.015 \times 10^{-2}.$$

Therefore, the quadrature error is several orders of magnitude smaller than the residual error associated with the spectral approximation and network optimization. Increasing the quadrature resolution beyond $K = 25$ produces no meaningful improvement in the numerical solution. Consequently, the value $K = 40$ adopted throughout the paper provides a reliable and conservative choice that ensures quadrature accuracy while avoiding unnecessary computational cost.

Table 10. Quadrature sensitivity results for Case V of Example 3.

K	J_K	$\frac{\lvert J_K - J_{\text{ref}}\rvert}{\lvert J_{\text{ref}}\rvert}$	$\lVert q_K - q_{\text{ref}}\rVert_{L^2}$	$\lVert z_K - z_{\text{ref}}\rVert_{L^2}$	$\lVert \mathcal{R}_q\rVert_{L^2}$	$\lVert \mathcal{R}_z\rVert_{L^2}$
5	0.464303909507	5.191×10^{-3}	1.304×10^{-2}	1.427×10^{-2}	1.405×10^{-2}	1.253×10^{-2}
9	0.467577683265	1.824×10^{-3}	3.254×10^{-3}	2.817×10^{-3}	1.145×10^{-2}	1.025×10^{-2}
13	0.466838114699	2.390×10^{-4}	6.070×10^{-4}	6.715×10^{-4}	1.125×10^{-2}	1.006×10^{-2}
17	0.466713862515	2.726×10^{-5}	5.352×10^{-5}	5.033×10^{-5}	1.135×10^{-2}	1.015×10^{-2}
25	0.466726593407	1.418×10^{-8}	2.203×10^{-8}	1.352×10^{-8}	1.135×10^{-2}	1.015×10^{-2}
40	0.466726603783	3.641×10^{-8}	4.569×10^{-8}	3.632×10^{-8}	1.135×10^{-2}	1.015×10^{-2}
60	0.466726586789	0	0	0	1.135×10^{-2}	1.015×10^{-2}

6. Conclusions

In this paper, a PI-CSNN framework was developed for solving distributed-order fractional optimal control problems. The proposed method combines the high-order accuracy of Chebyshev spectral approximations with a residual-based physics-informed training strategy, providing an accurate computational framework for distributed-order fractional systems. Numerical examples involving time-invariant and time-varying problems, including solutions with limited regularity, demonstrated that the PI-CSNN framework can produce highly accurate approximations of the state, adjoint, and control variables with systematic convergence behavior. The computed optimal cost values showed close agreement with previously reported numerical methods and benchmark results. Overall, the proposed PI-CSNN framework provides an accurate and adaptable approach to distributed-order fractional optimal control problems. Owing to its physics-informed spectral neural architecture, the framework may be extended to problems with box-constrained controls, multidimensional systems, and strongly nonlinear dynamics. However, such extensions introduce important computational and theoretical challenges, including the curse of dimensionality, efficient treatment of tensor-product approximation spaces, and increased complexity of nonlinear distributed-order operators. The numerical investigation of these extensions constitutes a promising direction for future research.

Use of Generative-AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU262887].

Conflict of interest

The author declares no conflict of interest.

References

1. F. Mainardi, *Fractional calculus and waves in linear viscoelasticity: an introduction to mathematical models*, New Jersey: World Scientific, 2022. https://doi.org/10.1142/p926
2. D. Baleanu, J. A. T. Machado, A. C. J. Luo, *Fractional dynamics and control*, New York: Springer, 2012. https://doi.org/10.1007/978-1-4614-0457-6
3. Z. Jiao, Y. Q. Chen, I. Podlubny, *Distributed-order dynamic systems: stability, simulation, applications and perspectives*, New York: Springer, 2012. https://doi.org/10.1007/978-1-4471-2852-6
4. T. M. Atanackovic, A generalized model for the uniaxial isothermal deformation of a viscoelastic body, *Acta Mech.*, **159** (2002), 77–86. https://doi.org/10.1007/BF01171449
5. A. V. Chechkin, R. Gorenflo, I. M. Sokolov, Retarding subdiffusion and accelerating superdiffusion governed by distributed-order fractional diffusion equations, *Phys. Rev. E*, **66** (2002), 046129. https://doi.org/10.1103/PhysRevE.66.046129
6. M. M. Meerschaert, H.-P. Scheffler, Stochastic model for ultraslow diffusion, *Stochastic Process. Appl.*, **116** (2006), 1215–1235. https://doi.org/10.1016/j.spa.2006.01.006
7. T. Sandev, A. V. Chechkin, N. Korabel, H. Kantz, I. M. Sokolov, R. Metzler, Distributed-order diffusion equations and multifractality: models and solutions, *Phys. Rev. E*, **92** (2015), 042117. https://doi.org/10.1103/PhysRevE.92.042117
8. H.-Y. Jian, T.-Z. Huang, X.-M. Gu, X.-L. Zhao, Y.-L. Zhao, Fast second-order implicit difference schemes for time distributed-order and Riesz space fractional diffusion-wave equations, *Comput. Math. Appl.*, **94** (2021), 136–154. https://doi.org/10.1016/j.camwa.2021.05.003
9. H. Ye, F. Liu, V. Anh, Compact difference scheme for distributed-order time-fractional diffusion-wave equation on bounded domains, *J. Comput. Phys.*, **298** (2015), 652–660. https://doi.org/10.1016/j.jcp.2015.06.025
10. G. H. Gao, A. A. Alikhanov, Z. Z. Sun, The temporal second order difference schemes based on the interpolation approximation for solving the time multi-term and distributed-order fractional sub-diffusion equations, *J. Sci. Comput.*, **73** (2017), 93–121. https://doi.org/10.1007/s10915-017-0407-x
11. M. L. Morgado, M. Rebelo, Numerical approximation of distributed order reaction-diffusion equations, *J. Comput. Appl. Math.*, **275** (2015), 216–227. https://doi.org/10.1016/j.cam.2014.07.029
12. M. Rebelo, M. L. Morgado, N. J. Ford, An implicit finite difference approximation for the solution of the diffusion equation with distributed order in time, *Electron. Trans. Numer. Anal.*, **44** (2015), 289–305.

13. W. Bu, A. Xiao, W. Zeng, Finite difference/finite element methods for distributed-order time fractional diffusion equations, *J. Sci. Comput.*, **72** (2017), 422–441. https://doi.org/10.1007/s10915-017-0360-8
14. W. Fan, F. Liu, A numerical method for solving the two-dimensional distributed order space-fractional diffusion equation on an irregular convex domain, *Appl. Math. Lett.*, **77** (2018), 114–121. https://doi.org/10.1016/j.aml.2017.10.005
15. M. A. Zaky, A Legendre collocation method for distributed-order fractional optimal control problems, *Nonlinear Dyn.*, **91** (2018), 2667–2681. https://doi.org/10.1007/s11071-017-4038-4
16. A. H. Bhrawy, S. S. Ezz-Eldien, A new Legendre operational technique for delay fractional optimal control problems, *Calcolo*, **53** (2016), 521–543. https://doi.org/10.1007/s10092-015-0160-1
17. E. Kharazmi, M. Zayernouri, G. E. Karniadakis, Petrov-Galerkin and spectral collocation methods for distributed order differential equations, *SIAM J. Sci. Comput.*, **39** (2017), A1003–A1037. https://doi.org/10.1137/16M1073121
18. A. Lischke, M. Zayernouri, G. E. Karniadakis, A Petrov-Galerkin spectral method of linear complexity for fractional multiterm ODEs on the half line, *SIAM J. Sci. Comput.*, **39** (2017), A922–A946. https://doi.org/10.1137/17M1113060
19. M. L. Morgado, M. Rebelo, L. L. Ferrás, N. J. Ford, Numerical solution for diffusion equations with distributed order in time using a Chebyshev collocation method, *Appl. Numer. Math.*, **114** (2017), 108–123. https://doi.org/10.1016/j.apnum.2016.11.001
20. J. Rashidinia, T. Eftekhari, K. Maleknejad, A novel operational vector for solving the general form of distributed order fractional differential equations in the time domain based on the second kind Chebyshev wavelets, *Numer. Algorithms*, **88** (2021), 1617–1639. https://doi.org/10.1007/s11075-021-01088-8
21. L. Wei, Y. Chen, F. Qin, J. Huang, P. Li, The discontinuous Galerkin method for solving the two-dimensional transport equation with distributed-order memory kernel on triangular meshes, *Appl. Numer. Math.*, **228** (2026), 57–70. https://doi.org/10.1016/j.apnum.2026.05.009
22. S. Haykin, *Neural networks: a comprehensive foundation*, New Jersey: Prentice Hall, 1998.
23. J. M. Zurada, *Introduction to artificial neural systems*, St. Paul: West Publishing Company, 1992.
24. A. M. Neves, W. Großmann, J. A. Puskiel, J. Warfsmann, V. R. Hosseini, M. Passing, et al., Kinetic-model identification in metal-hydride reactions using neural network autoencoder surrogate models, *Energy AI*, **22** (2025), 100659. https://doi.org/10.1016/j.egyai.2025.100659
25. V. R. Hosseini, A. A. Mehrizi, A. Gungor, H. H. Afrouzi, Application of a physics-informed neural network to solve the steady-state Bratu equation arising from solid biofuel combustion theory, *Fuel*, **332** (2023), 125908. https://doi.org/10.1016/j.fuel.2022.125908
26. C. Leake, H. Johnston, L. Smith, D. Mortari, Analytically embedding differential equation constraints into least squares support vector machines using the theory of functional connections, *Mach. Learn. Knowl. Extr.*, **1** (2019), 1058–1083. https://doi.org/10.3390/make1040060
27. A. A. Aghaei, K. Parand, Hyperparameter optimization of orthogonal functions in the numerical solution of differential equations, *Math. Methods Appl. Sci.*, **48** (2025), 605–625. https://doi.org/10.1002/mma.10346

28. S. Mall, S. Chakraverty, Chebyshev neural network based model for solving Lane-Emden type equations, *Appl. Math. Comput.*, **247** (2014), 100–114. https://doi.org/10.1016/j.amc.2014.08.085
29. W. H. Luo, X. M. Gu, B. Carpentieri, J. Guo, A Bernoulli-barycentric rational matrix collocation method with preconditioning for a class of evolutionary PDEs, *Numer. Linear Algebra Appl.*, **32** (2025), e70007. https://doi.org/10.1002/nla.70007
30. X. M. Gu, H. W. Sun, Y. L. Zhao, X. Zheng, An implicit difference scheme for time-fractional diffusion equations with a time-invariant type variable order, *Appl. Math. Lett.*, **120** (2021), 107270. https://doi.org/10.1016/j.aml.2021.107270
31. X. M. Gu, S. L. Wu, A parallel-in-time iterative algorithm for Volterra partial integro-differential problems with weakly singular kernel, *J. Comput. Phys.*, **417** (2020), 109576. https://doi.org/10.1016/j.jcp.2020.109576
32. J. Zhou, X. M. Gu, Y. L. Zhao, H. Li, A fast compact difference scheme with unequal time-steps for the tempered time-fractional Black-Scholes model, *Int. J. Comput. Math.*, **101** (2024), 989–1011. https://doi.org/10.1080/00207160.2023.2254412
33. T. Liu, B. Ding, A radial basis function neural network approach for solving a diffusion partial differential equation efficiently, *Appl. Math. Comput.*, **509** (2026), 129651. https://doi.org/10.1016/j.amc.2025.129651
34. M. Vellappandi, S. Lee, Physics-informed neural fractional differential equations, *Appl. Math. Model.*, **145** (2025), 116127. https://doi.org/10.1016/j.apm.2025.116127
35. A. Lachin, M. A. Abdelkawy, S. Sathasivam, Spectral collocation method for solving nonlinear Riesz distributed-order fractional differential equations, *Mathematics*, **13** (2025), 3425. https://doi.org/10.3390/math13213425
36. R. Balamurugan, I. Khan, M. Bakouri, T. Alqahtani, Physics-informed neural network approach to unsteady fractional flow in a vertical coaxial annulus with thermal effects and magneto-hall interaction, *Results Eng.*, **29** (2026), 109965. https://doi.org/10.1016/j.rineng.2026.109965
37. A. Alizadeh, S. Effati, An iterative approach for solving fractional optimal control problems, *J. Vib. Control*, **24** (2018), 18–36. https://doi.org/10.1177/1077546316633391
38. S. Soradi Zeid, M. Yousefi, M. Yousefi, Approximated solutions of linear quadratic fractional optimal control problems, *J. Appl. Math. Stat. Inform.*, **12** (2016), 83–94. https://doi.org/10.1515/jamsi-2016-0010
39. P. Sahu, S. Saha Ray, Comparison on wavelets techniques for solving fractional optimal control problems, *J. Vib. Control*, **24** (2018), 1185–1201. https://doi.org/10.1177/1077546316659611
40. E. H. Doha, A. H. Bhrawy, D. Baleanu, R. M. Hafez, M. A. Abdelkawy, An efficient numerical scheme based on the shifted orthonormal Jacobi polynomials for solving fractional optimal control problems, *Adv. Differ. Equ.*, **2015** (2015), 15. https://doi.org/10.1186/s13662-014-0344-z

AIMS Press