Networks and Heterogeneous Media

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

NHM, 21(7): 2556–2574.
DOI: 10.3934/nhm.2026095
Received: 29 August 2026
Revised: 17 October 2026
Accepted: 28 October 2026
Published: 30 November 2026

Research article

Residual bounds for singularity-corrected fractional physics-informed neural networks for the two-dimensional Allen–Cahn equation

Lijie Liu and Leilei Wei*

School of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China

* **Correspondence:** Email: leileiwei09@gmail.com, leileiwei@nmu.edu.cn.

Abstract: We developed a singularity-corrected fractional physics-informed neural network (fPINN) for the two-dimensional time-fractional Allen–Cahn equation. The trial function enforces the initial and homogeneous Dirichlet data and includes two fractional powers whose coefficients are determined by the initial data and forcing. The method evaluates the Caputo derivative of these terms analytically and applies the nonuniform L1 formula to the neural remainder. For the corrected L1 scheme with a five-point spatial stencil, the discrete residual gives a computable maximum-norm bound on the error relative to the discrete solution. The comparison uses the signs of the L1 coefficients and the one-sided Lipschitz property of the cubic reaction under an explicit time-step restriction. With additional regularity, temporal and spatial consistency estimates give an error bound at the nodes relative to the exact partial differential equation (PDE) solution. Over three random initializations, the correction reduced the mean relative L^2 test error by 74.6% for the Mixed method and 57.9% for the Localized method compared with the plain fPINN. In the unforced test, the maximum discrete error bound fell below 0.02 after one or two spatial enrichment stages in each of the three runs.

Keywords: fractional physics-informed neural networks; time-fractional Allen–Cahn equation; singularity correction; residual error bounds; nonuniform L1 method

1. Introduction

The Allen–Cahn equation describes antiphase boundary motion and domain coarsening [1]. Replacing the first-order time derivative with a Caputo derivative introduces memory into the phase-field model [2]. Sakamoto and Yamamoto [3] developed the theory of initial-boundary value problems for linear fractional diffusion. For the time-fractional Allen–Cahn equation, Du et al. [4] proved well-posedness, described the limited solution regularity, and derived error estimates.

The time derivative can be singular near $t = 0$ for smooth spatial data, limiting the accuracy of time discretizations. Ji et al. [5] constructed maximum-principle-preserving L1 schemes for the time-

fractional Allen–Cahn equation, and Liao et al. [6] obtained the same property for a second-order nonuniform scheme. Tang et al. [7] derived an integral energy dissipation law and developed an energy-stable scheme. A compatible energy law for the variable-step L1 scheme was established in [8]. Mao et al. [9] used a lattice Boltzmann method to evaluate the fractional operator and study memory-dependent interface motion. Zhang et al. [10] developed a compact finite-difference scheme for Caputo–Fabrizio diffusion. For convection–diffusion–reaction equations with a Caputo derivative, Zhang et al. [11] combined high-order time discretization with sixth-order finite differences in space. Wei et al. [12] used a Grünwald-type time approximation and a discontinuous Galerkin discretization on triangular meshes for two-dimensional transport with distributed-order memory.

The L1 formula approximates the Caputo history using a piecewise linear interpolant. Lin and Xu [13] analyzed finite-difference and spectral approximations for time-fractional diffusion. Initial singularities reduce the temporal accuracy on uniform meshes [14]. Graded meshes resolve the initial layer by placing more time nodes near $t = 0$ [15, 16]. They have also been used in discontinuous Galerkin approximations of natural convection with tempered fractional constitutive relations [17]. Error estimates for nonuniform L1 formulas [18] and discrete fractional Gronwall inequalities [19] provide tools for the analysis of variable time steps. Starting corrections account for noninteger powers in fractional discretizations [20]. A neural method that uses these history formulas inherits their discretization error in addition to the error of the neural representation.

Physics-informed neural networks (PINNs) approximate solutions of partial differential equations (PDEs) by minimizing equation residuals and mismatches in the data [21,22]. One application is the two-dimensional sine–Gordon equation [23]. Fractional PINNs (fPINNs) combine neural representations with discrete nonlocal operators [24]. Monte Carlo variants evaluate fractional constraints by sampling in high-dimensional problems [25]. For fractional phase-field equations, Kang et al. [26] combined the L1 formula with adaptive sampling and optimization. Gradient-flow and neural-tangent-kernel analyses relate imbalances between loss terms to PINN training [27, 28]. Adaptive loss weighting [29], gradient-enhanced residual losses [30], and causality-aware training [31] address different aspects of the optimization problem. DeepXDE implements residual-based point enrichment [32], and Wu et al. [33] compared residual-based and other sampling strategies. Enforcing the initial and boundary data in the trial function removes the corresponding loss terms [34].

SC-PIELM uses a fractional time input and a high-order history formula in a physics-informed extreme learning machine [35]. DSC-PINN separates singular and regular solution components [36]. For phase-field equations, ASA-fPINNs use graded time grids and accelerated history evaluation [37]. In the method below, the problem data determine the lifting coefficients, and the corrected history formula is used in a nonlinear residual comparison estimate.

Residual analysis for PINNs includes convergence results for linear elliptic and parabolic problems [38] and generalization estimates based on PDE stability [39]. De Ryck and Mishra [40] review approximation, sampling, and training errors. For fractional parabolic equations, pointwise-in-time a posteriori estimates control errors by means of stability bounds [41, 42].

The corrected L1 scheme retains the signs of the history coefficients. Together with $(a^3 - a - b^3 + b)(a - b) \geq -|a - b|^2$, these signs give a maximum-norm comparison recurrence in which the prescribed lifting correction cancels. The recurrence bounds the error relative to the exact solution of the corrected discrete equations and also compares two approximations with nonzero residuals. Under additional regularity, temporal and spatial consistency estimates give a bound for the PDE error at the grid nodes.

In the unforced test, the discrete bound determines when to enrich the spatial collocation set and when to stop training.

2. Model and initial expansion

Let $\Omega = (0,1)^2$ and $T > 0$. We consider

$$\begin{aligned} {}^{C}D_t^{\alpha} u - \varepsilon^2 \Delta u + F(u) &= f \quad \text{in } \Omega \times (0,T], \\ u &= 0 \quad \text{on } \partial\Omega \times [0,T], \\ u(\cdot,0) &= g \quad \text{in } \Omega, \end{aligned} \tag{2.1}$$

where Δ is the spatial Laplacian, $\varepsilon > 0$, and $F(u) = u^3 - u$. Here $F = \Psi'$ with $\Psi(u) = \frac{1}{4}(u^2-1)^2$. Diffusion smooths spatial variations, and Ψ has minima at $u = \pm 1$. The Caputo derivative is defined by

$${}^{C}D_t^{\alpha} v(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} v'(s)\, ds, \qquad 0 < \alpha < 1.$$

Assume that u is absolutely continuous in time with values in $C(\overline{\Omega})$. We determine the lifting coefficients from the expansions

$$\begin{aligned} f(t) &= f_0 + f_\alpha t^{\alpha} + o(t^{\alpha}), \\ u(t) &= g + s_1 t^{\alpha} + s_2 t^{2\alpha} + o(t^{2\alpha}), \end{aligned} \tag{2.2}$$

as $t \downarrow 0$. The expansions for f and u are in $C(\overline{\Omega})$ and $C^2(\overline{\Omega})$, respectively, so the solution expansion can be used in Δu in the Volterra equation

$$u(t) = g + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \big[f(s) + \varepsilon^2 \Delta u(s) - F(u(s)) \big]\, ds.$$

We match powers in this integral equation to determine s_1 and s_2. The analytic history correction uses the identity

$${}^{C}D_t^{\alpha} t^{\sigma} = \frac{\Gamma(1+\sigma)}{\Gamma(1+\sigma-\alpha)} t^{\sigma-\alpha}, \qquad \sigma > 0. \tag{2.3}$$

The constant and t^{α} coefficients of the Volterra integrand give

$$\begin{aligned} s_1 &= \frac{f_0 + \varepsilon^2 \Delta g - g^3 + g}{\Gamma(1+\alpha)}, \\ s_2 &= \frac{\Gamma(1+\alpha)}{\Gamma(1+2\alpha)} \big[f_\alpha + \varepsilon^2 \Delta s_1 - (3g^2 - 1) s_1 \big]. \end{aligned} \tag{2.4}$$

The fractional integral in the Volterra formulation acts on the powers by $I^{\alpha} t^{\beta} = \Gamma(1+\beta) t^{\alpha+\beta} / \Gamma(1+\alpha+\beta)$ for $\beta \geq 0$. For the unforced problem, $f_0 = f_\alpha = 0$.

Define

$$S(x,y,t) = s_1(x,y) t^{\alpha} + s_2(x,y) t^{2\alpha}, \qquad v = u - g - S.$$

The temporal consistency estimate uses the following regularity of v:

$$v \in C^2([0,T]; L^{\infty}(\Omega)), \quad M_t = \sup_{0 \leq t \leq T} \left\| \partial_t^2 v(t) \right\|_{\infty} < \infty. \tag{2.5}$$

A forcing term proportional to t can generate a $t^{1+\alpha}$ term in u. If present, this term must be included in the lifting to obtain a C^2 remainder. Additional powers can be treated by Eq (2.3). Graded meshes provide a setting for analysis under weaker temporal regularity.

The compatibility condition for the trial function is $g = s_1 = s_2 = 0$ on $\partial\Omega$. For $g \in C^6(\overline{\Omega})$, $f_0 \in C^4(\overline{\Omega})$, and $f_\alpha \in C^2(\overline{\Omega})$, Eq (2.4) gives $s_1 \in C^4(\overline{\Omega})$ and $s_2 \in C^2(\overline{\Omega})$, with classical Laplacians. The boundary traces are evaluated from Eq (2.4). Incompatible or time-dependent Dirichlet data require an additional boundary lifting.

3. Singularity-corrected fractional neural residual

3.1. Trial function and exact data enforcement

Let $\rho(x, y) = 16x(1 - x)y(1 - y)$ and let $\mathcal{N}_\theta$ be a feedforward network. We set

$$\begin{aligned} u_\theta(x, y, t) &= g(x, y) + S(x, y, t) + v_\theta(x, y, t), \\ v_\theta(x, y, t) &= t^2\rho(x, y)\mathcal{N}_\theta(2x - 1, 2y - 1, 2t/T - 1). \end{aligned} \tag{3.1}$$

The hidden layers use tanh activations, and the output layer is linear. The factor ρ vanishes on $\partial\Omega$ and is normalized by $\max_{\overline{\Omega}} \rho = 1$. With compatible g, s_1, s_2, Eq (3.1) enforces the initial and boundary conditions. The experiments use $2/3 < \alpha < 1$, for which $t^{3\alpha} = o(t^2)$. In this range, $v = o(t^{2\alpha})$ and Eq (2.5) give $v(0) = \partial_t v(0) = 0$ and $v(t) = O(t^2)$ in $L^\infty(\Omega)$. The time factor t^2 in Eq (3.1) reflects this behavior. For each finite parameter vector θ, $\partial_t^2 v_\theta$ is bounded on $\overline{\Omega} \times [0, T]$.

The coefficients s_1, s_2 are fixed during training. The manufactured tests use their equivalent closed-form expressions.

3.2. Nonuniform L1 formula

We approximate the history of the remainder by the nonuniform L1 formula, which uses piecewise linear interpolation [13, 18]. For $0 = t_0 < t_1 < \cdots < t_N = T$, let $\tau_j = t_j - t_{j-1}$ and $\tau_{\max} = \max_j \tau_j$. The formula is

$$(\delta_\tau^\alpha w)^n = \sum_{j=1}^{n} a_{nj}(w^j - w^{j-1}), \quad a_{nj} = \frac{(t_n - t_{j-1})^{1-\alpha} - (t_n - t_j)^{1-\alpha}}{\Gamma(2 - \alpha)\tau_j}. \tag{3.2}$$

Let $D \in \mathbb{R}^{N\times(N+1)}$ map the nodal values $(w^0, \ldots, w^N)$ to the L1 derivatives in Eq (3.2). Its nonzero entries are

$$D_{n0} = -a_{n1}, \quad D_{nj} = a_{nj} - a_{n,j+1}\ (1 \le j < n), \quad D_{nn} = a_{nn}. \tag{3.3}$$

The implementation uses `log1p` and `expm1` to evaluate differences of nearby fractional powers. Define the corrected derivative and additive correction by

$$\begin{aligned} (\widehat{\delta_\tau^\alpha} u_\theta)^n &= {}^C D_t^\alpha S(t_n) + (\delta_\tau^\alpha v_\theta)^n, \\ C_S^n &= {}^C D_t^\alpha S(t_n) - (\delta_\tau^\alpha S)^n. \end{aligned} \tag{3.4}$$

Since g is independent of time, the corrected derivative is $(\delta_\tau^\alpha u_\theta)^n + C_S^n$ with the original L1 coefficients. The term ${}^C D_t^\alpha S$ is evaluated by Eq (2.3).

3.3. Collocation points, loss, and optimization

Let $\{z_i\}_{i=1}^{M} \subset \Omega$ be fixed spatial collocation points. We evaluate the network directly at each (z_i, t_j) for $0 \le j \le N$. The training residual is

$$R_{\theta,i}^{n} = (\widehat{\delta_\tau^\alpha} u_\theta)^n(z_i) - \varepsilon^2 \Delta u_\theta(z_i, t_n) + u_\theta(z_i, t_n)^3 - u_\theta(z_i, t_n) - f(z_i, t_n).$$

Spatial derivatives are evaluated by automatic differentiation. The loss is

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \sum_{n=1}^{N} \frac{\tau_n}{T} |R_{\theta,i}^{n}|^2. \tag{3.5}$$

The weights τ_n/T account for the time interval lengths in the average of the squared residual. We minimize this loss with Adam followed by L-BFGS with a strong-Wolfe line search.

Algorithm 1. Corrected fPINN and residual evaluation

Step	Procedure
1	Determine s_1, s_2 from Eq (2.4) and check boundary compatibility.
2	Select the time mesh and spatial collocation points, and then construct D, S, and ${}^{C}D_t^\alpha S$.
3	Initialize Eq (3.1) and minimize Eq (3.5) with the prescribed optimizer budget.
4	Evaluate the network at every space-time node of the Cartesian audit grid and compute Eq (4.5).
5	Set $\eta_n = \|\mathcal{R}^n\|_{\infty,h}$ and compute $z = B^{-1}\eta$ using Eq (4.8).
6	Stop when the bound meets the prescribed discrete tolerance. Otherwise, add spatial collocation points and continue within the specified budget.

The manufactured-solution tests use fixed optimizer budgets. For the unforced test, the rule for adding spatial points is given in Section 5.6. All experiments evaluate the history directly. The history product costs $O(MN^2)$ per loss evaluation, in addition to the network and automatic-differentiation costs. The neural values and history matrix require $O(MN)$ and $O(N^2)$ storage, respectively. Given the residual, evaluation of the bound costs $O(N^2)$. Sum-of-exponentials compression [43] can reduce the history cost. Its approximation error would contribute to the residual assessment.

4. Consistency and residual comparison

4.1. L1 coefficient signs and temporal consistency

Lemma 4.1. *On any strictly increasing time mesh, $a_{nj} > 0$ and $a_{n,j+1} \ge a_{nj}$ for $j < n$. Thus $D_{n0} < 0$, $D_{nj} \le 0$ for $1 \le j < n$, and each row of D sums to zero.*

Proof. Each a_{nj} averages the positive, increasing kernel $(t_n - s)^{-\alpha}/\Gamma(1-\alpha)$ over $[t_{j-1}, t_j]$. The successive averages increase with j. The signs and row sums of D follow from Eq (3.3). □

Theorem 4.2 (Consistency of the corrected history)**.** *On a strictly increasing time mesh, let $u = g + S + v$ with v satisfying Eq* (2.5)*. Then, at every positive time node,*

$$\left\| (\delta_\tau^\alpha u)^n + C_S^n - {}^{C}D_t^\alpha u(t_n) \right\|_\infty \le \frac{M_t}{2\Gamma(2-\alpha)} \tau_{\max}^{2-\alpha}. \tag{4.1}$$

At t_n, the estimate also holds with $\tau_{\max,n} = \max_{1 \le j \le n} \tau_j$ in place of $\tau_{\max}$.

Proof. The error in Eq (4.1) equals $\left\|(\delta_\tau^\alpha v)^n - {}^C D_t^\alpha v(t_n)\right\|_\infty$. Let $I_\tau v$ denote the piecewise linear interpolant of v and set $e = v - I_\tau v$. For $s \in [t_{j-1}, t_j]$, the interpolation estimate gives

$$\|e(s)\|_\infty \le \tfrac{1}{2} M_t (s - t_{j-1})(t_j - s) \le \tfrac{1}{2} M_t \tau_{\max} \min\{\tau_{\max}, t_n - s\}.$$

Integrate by parts on each time interval. The endpoint terms at the nodes vanish since e vanishes there. At $s = t_n$, the weighted endpoint term is $O((t_n - s)^{1-\alpha})$ and tends to zero. Thus

$$\left\|{}^C D_t^\alpha v(t_n) - (\delta_\tau^\alpha v)^n\right\|_\infty \le \frac{\alpha}{\Gamma(1-\alpha)} \int_0^{t_n} (t_n - s)^{-1-\alpha} \|e(s)\|_\infty \, ds.$$

With $r = t_n - s$, extend the integral to $(0, \infty)$ and split it at $\tau_{\max}$ to obtain

$$\frac{\alpha M_t}{2\Gamma(1-\alpha)} \left[\tau_{\max} \frac{\tau_{\max}^{1-\alpha}}{1-\alpha} + \tau_{\max}^2 \frac{\tau_{\max}^{-\alpha}}{\alpha} \right] = \frac{M_t \tau_{\max}^{2-\alpha}}{2\Gamma(2-\alpha)}.$$

□

The uncorrected L1 derivative of t^α on a uniform mesh has first-node error

$$(\delta_\tau^\alpha t^\alpha)^1 - {}^C D_t^\alpha t^\alpha(t_1) = \frac{1}{\Gamma(2-\alpha)} - \Gamma(1+\alpha), \tag{4.2}$$

which is independent of the step size. Its contribution to the time-stepping error is governed by the inverse discrete history operator.

4.2. Corrected discrete problem

Let $h = 1/K$, and let Ω_h be the interior nodes of the uniform Cartesian grid. For grid functions with homogeneous boundary values, define

$$(\Delta_h w)_{ij} = h^{-2}(w_{i+1,j} + w_{i-1,j} + w_{i,j+1} + w_{i,j-1} - 4w_{ij}). \tag{4.3}$$

A subscript h denotes restriction to the spatial grid and F acts componentwise. Let U_h satisfy the corrected discrete equations

$$(\delta_\tau^\alpha U_h)^n + C_{S,h}^n - \varepsilon^2 \Delta_h U_h^n + F(U_h^n) = f_h^n, \qquad U_h^0 = g_h. \tag{4.4}$$

For any candidate sequence W^n with the same initial and boundary values, define

$$\mathcal{R}^n = (\delta_\tau^\alpha W)^n + C_{S,h}^n - \varepsilon^2 \Delta_h W^n + F(W^n) - f_h^n, \quad \eta_n = \|\mathcal{R}^n\|_{\infty,h}. \tag{4.5}$$

The norm $\|\cdot\|_{\infty,h}$ is the maximum over the interior audit nodes. For a neural approximation, W^n contains the network values at those nodes. The audit residual uses the finite-difference Laplacian in Eq (4.4). The training residual uses automatic differentiation for the spatial derivatives.

Assume

$$a_{nn} = \frac{\tau_n^{-\alpha}}{\Gamma(2-\alpha)} > 1 \qquad (1 \le n \le N). \tag{4.6}$$

Equivalently, $\tau_n < \Gamma(2-\alpha)^{-1/\alpha}$. Define the lower triangular matrix $B \in \mathbb{R}^{N \times N}$ by

$$B_{nn} = a_{nn} - 1, \quad B_{nj} = D_{nj} \ (1 \le j < n), \quad B_{nj} = 0 \ (j > n).$$

The matrix B uses the positive time levels of D. The vectors $\eta, z \in \mathbb{R}^N$ use the same indexing, and their inequalities are componentwise.

Proposition 4.3 (Well-posedness and positivity)**.** *Under Eq* (4.6)*, each nonlinear step in Eq* (4.4) *has a unique solution, and B^{-1} exists and is entrywise nonnegative.*

Proof. Collecting the known history terms gives the step

$$[(a_{nn} - 1)I - \varepsilon^2 \Delta_h]w + w^{\circ 3} = b.$$

It is the Euler equation of the coercive strictly convex function

$$\tfrac{1}{2}\left([(a_{nn} - 1)I - \varepsilon^2 \Delta_h]w, w\right) + \tfrac{1}{4}\sum_{ij} w_{ij}^4 - (b, w)\,.$$

The positive definiteness of $-\Delta_h$ and Eq (4.6) make the functional strictly convex and coercive. Its unique minimizer solves the nonlinear step. By Lemma 4.1, B has a positive diagonal and nonpositive entries below it. Forward substitution in $Bz = q$ gives $z \geq 0$ for $q \geq 0$, so $B^{-1} \geq 0$ entrywise. □

Theorem 4.4 (Residual comparison certificate)**.** *Let $0 < \alpha < 1$ and assume Eq* (4.6) *holds. For any prescribed correction $C_{S,h}^n$ in Eq* (4.4)*, let the real-valued sequence W satisfy the prescribed initial and boundary conditions. Then*

$$\left\|W^n - U_h^n\right\|_{\infty,h} \leq z_n, \qquad z = B^{-1}\eta. \tag{4.7}$$

The entries of z follow from the residual norms by forward substitution according to

$$z_n = \frac{\eta_n + \sum_{j=1}^{n-1}(a_{n,j+1} - a_{nj})z_j}{a_{nn} - 1}. \tag{4.8}$$

Proof. Set $e^n = W^n - U_h^n$ and $E_n = \|e^n\|_{\infty,h}$. The correction $C_{S,h}^n$ cancels on subtracting Eq (4.4) from Eq (4.5). If $E_n > 0$, take an interior node where $|e^n| = E_n$ and multiply the error equation there by $\operatorname{sign}(e^n)$. The signed diffusion term is nonnegative at this extremum because the boundary error is zero. The reaction difference satisfies

$$F(W) - F(U_h) = (W^2 + WU_h + U_h^2 - 1)e, \qquad W^2 + WU_h + U_h^2 \geq 0.$$

Thus the signed reaction term is bounded below by $-E_n$. Using the signs of the history coefficients yields

$$(a_{nn} - 1)E_n \leq \eta_n + \sum_{j=1}^{n-1}(a_{n,j+1} - a_{nj})E_j.$$

This inequality also holds when $E_n = 0$. Hence $BE \leq \eta$, and $B^{-1} \geq 0$ gives the result. □

Corollary 4.5 (Two inexact sequences)**.** *Assume Eq* (4.6) *holds. Let W and V satisfy the same prescribed initial and boundary conditions on the same grid. Let η^W and η^V be their residual norms from Eq* (4.5) *for the same correction and forcing. Then*

$$\|W^n - V^n\|_{\infty,h} \leq \left[B^{-1}(\eta^W + \eta^V)\right]_n. \tag{4.9}$$

Proof. Apply Theorem 4.4 to each sequence relative to the unique U_h and use the triangle inequality. □

For a Newton-computed reference, Eq (4.9) includes its residual. If W has zero boundary error and initial error $E_0 = \left\|W^0 - U_h^0\right\|_{\infty,h}$, the bound vector in Eq (4.7) becomes $B^{-1}(\eta - D_{\cdot 0}E_0)$, where $D_{\cdot 0} = (D_{10}, \ldots, D_{N0})^\top$. The initial-error contribution $-D_{\cdot 0}E_0$ is nonnegative by Lemma 4.1.

4.3. Error relative to the PDE solution at grid nodes

Corollary 4.6 (Nodal PDE error bound)**.** *Assume the hypotheses of Theorem 4.4 and Eq* (2.5)*. Let u solve Eq* (2.1) *with spatial derivatives through order four continuous on* $\overline{\Omega}$ *and uniformly bounded for* $t \in [0, T]$*. Set*

$$M_4 = \sup_{t\in[0,T]} \left(\|u_{xxxx}(t)\|_\infty + \left\|u_{yyyy}(t)\right\|_\infty \right),$$
$$Q = \frac{M_t \tau_{\max}^{2-\alpha}}{2\Gamma(2-\alpha)} + \frac{\varepsilon^2 h^2 M_4}{12}.$$

Let $\mathbf{1} \in \mathbb{R}^N$ *be the vector of ones and* u_h^n *the exact nodal samples. Then*

$$\left\|W^n - u_h^n\right\|_{\infty,h} \le [B^{-1}(\eta + Q\mathbf{1})]_n. \tag{4.10}$$

Proof. The residual of u_h in Eq (4.5) consists of the temporal and spatial consistency defects. By Theorem 4.2 and the central-difference remainder, their sum has a maximum norm of at most Q. Apply Corollary 4.5 to W and u_h. □

5. Numerical experiments

5.1. Implementation and evaluation

Unless stated otherwise, we use $N = 32$, $t_n = Tn/N$, and an audit spacing of $h = 1/32$. All experimental time grids satisfy Eq (4.6). The smallest margin is $\min_n(a_{nn} - 1) = 6.7600$ for $\alpha = 0.7$, $T = 1$, and $N = 16$ on the uniform mesh.

The manufactured-solution test set $\mathcal{T}$ consists of 39×39 interior spatial points and 80 temporal midpoints $T(k + 1/2)/80$, $0 \le k < 80$. We reserve this set for error evaluation. The sampled errors are

$$E_2 = \left(\frac{\sum_{(z,t)\in\mathcal{T}} |u_\theta(z,t) - u(z,t)|^2}{\sum_{(z,t)\in\mathcal{T}} |u(z,t)|^2} \right)^{1/2}, \quad E_\infty = \max_{(z,t)\in\mathcal{T}} |u_\theta(z,t) - u(z,t)|. \tag{5.1}$$

The early-time error is $E_{\text{early}} = \max_{(z,t)\in\mathcal{T},\, t\le 0.1T} |u_\theta(z,t) - u(z,t)|$. We compute means and sample standard deviations from all three initialization seeds.

The reference solver uses fully implicit L1 time stepping and the five-point Laplacian, with damped Newton iteration at each time level. Newton iteration stops when the maximum norm of the step residual is at most 10^{-11}. Backtracking starts with a unit step and halves it until the residual ℓ^2 norm decreases or the damping reaches $1/128$. The iteration cap is 30, and a step is rejected if its final maximum residual exceeds 10^{-9}. The final residual is recorded for Eq (4.9). We evaluate the residual bounds in standard double precision. The reference refinement and consistency calculations are carried out separately from network training.

In Table 1, Feature uses both a fractional time input and a leading t^α factor. Lifted and Corrected share the same trial function and differ in the history formula. A comparison with Plain also reflects the prescribed initial expansion in the lifted models. We assess Plain, Feature, and Lifted using the uncorrected discrete equations ($C_S = 0$) and Corrected using Eq (4.4). Within each manufactured test, the four models use the same exact solution for error evaluation.

Table 1. Ablation models with a common network architecture and optimizer budget. The time inputs are rescaled to $[-1, 1]$ before they enter the network.

Label	Trial function	Time input	History evaluation
Plain	$g + t\rho\mathcal{N}$	t/T	L1 on full output
Feature	$g + t^{\alpha}\rho\mathcal{N}$	$(t/T)^{\alpha}$	L1 on full output
Lifted	$g + S + t^2\rho\mathcal{N}$	t/T	L1 on full output
Corrected	$g + S + t^2\rho\mathcal{N}$	t/T	exact ${}^{C}D_t^{\alpha}S$ plus L1 on remainder

5.2. *Two manufactured solutions*

Let $\phi_{pq}(x, y) = \sin(p\pi x)\sin(q\pi y)$ and

$$H(x, y) = 16x(1 - x)y(1 - y)\exp[-18((x - 0.35)^2 + (y - 0.6)^2)].$$

Both tests use $T = 1$, $\alpha = 0.7$, and

$$u_*(x, y, t) = g + s_1 t^{\alpha} + s_2 t^{2\alpha} + r_2 t^2 + r_3 t^3, \tag{5.2}$$

with the coefficients in Table 2. The two values of ε give different diffusion strengths for these smooth-solution tests. Mixed uses trigonometric modes with different time factors. Localized includes the Gaussian-modulated profile H. The remainder $r_2 t^2 + r_3 t^3$ in each test satisfies Eq (2.5).

Table 2. Coefficients of the manufactured solutions in Eq (5.2).

Case	ε	g	s_1	s_2	r_2	r_3
Mixed	0.1	$0.2\phi_{11}$	$0.25\phi_{12}$	$-0.12\phi_{21}$	$0.15\phi_{11}$	$0.08\phi_{22}$
Localized	0.05	$0.6\phi_{11}$	$-0.25H$	$0.1\phi_{22}$	$0.2H$	$-0.1\phi_{12}$

Substituting u_* in Eq (2.1) gives the forcing

$$f = \sum_{\sigma\in\{\alpha,2\alpha,2,3\}} \frac{\Gamma(1 + \sigma)}{\Gamma(1 + \sigma - \alpha)} t^{\sigma-\alpha} c_{\sigma} - \varepsilon^2 \Delta u_* + u_*^3 - u_*,$$

where $(c_{\alpha}, c_{2\alpha}, c_2, c_3) = (s_1, s_2, r_2, r_3)$. We evaluate the spatial derivatives of u_* before training. The first two source coefficients are

$$f_0 = \Gamma(1 + \alpha)s_1 - \varepsilon^2 \Delta g + g^3 - g,$$
$$f_{\alpha} = \frac{\Gamma(1 + 2\alpha)}{\Gamma(1 + \alpha)} s_2 - \varepsilon^2 \Delta s_1 + (3g^2 - 1)s_1.$$

Substitution into Eq (2.4) recovers the prescribed s_1 and s_2 from the initial data and forcing.

5.3. *Ablation results and initialization variability*

Table 3 gives mean E_2 values of 4.328×10^{-3} for Mixed and 2.321×10^{-3} for Localized with Corrected. The corresponding Plain values are 1.701×10^{-2} and 5.517×10^{-3}, giving reductions of

approximately 74.6% and 57.9%. With the common lifted trial function, the analytic history correction reduces the mean E_2 by 27.4% and 20.2% relative to Lifted at the same optimizer budget.

Table 3. Manufactured-solution errors at $N = 32$, $M = 128$, and $\alpha = 0.7$. The E_2 and E_∞ columns report means ± sample standard deviations over three initializations; the other columns report means. Errors are multiplied by 10^3 and losses by 10^6.

Case	Method	$10^3 E_2$	$10^3 E_\infty$	$10^3 E_{\text{early}}$	$10^6 \mathcal{L}$
Mixed	Plain	17.007 ± 0.612	8.983 ± 0.190	8.024	23.817
Mixed	Feature	14.616 ± 0.233	7.659 ± 0.206	5.192	4.727
Mixed	Lifted	5.960 ± 0.306	4.582 ± 0.304	0.277	23.008
Mixed	Corrected	4.328 ± 0.251	3.442 ± 0.271	0.160	1.205
Localized	Plain	5.517 ± 0.262	7.415 ± 0.151	7.415	7.169
Localized	Feature	5.127 ± 0.102	6.501 ± 0.685	4.762	3.994
Localized	Lifted	2.910 ± 0.161	3.747 ± 0.377	0.226	5.327
Localized	Corrected	2.321 ± 0.129	3.498 ± 0.389	0.142	0.573

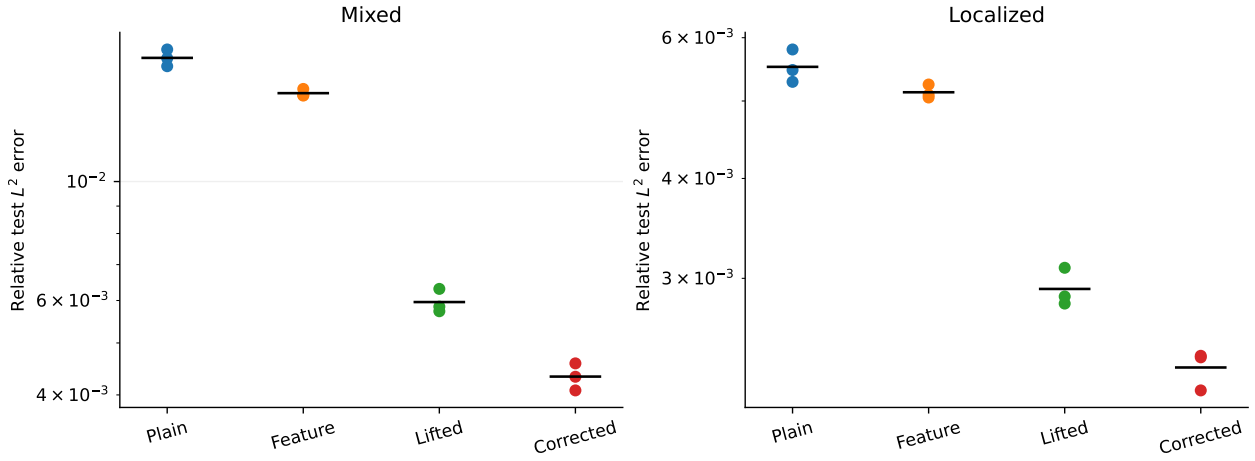


Figure 1. Test errors for three initialization seeds. Points denote individual initializations and black bars mark arithmetic means. The vertical axis is logarithmic.

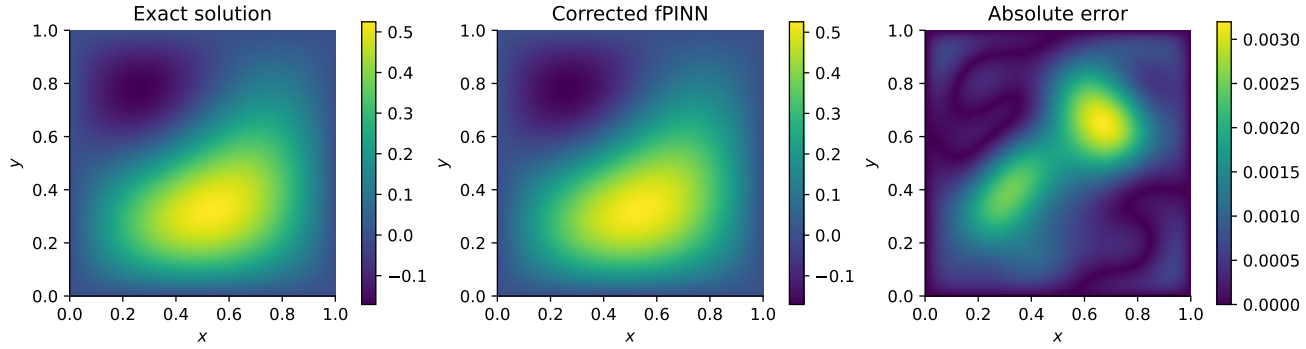


Figure 2. Exact solution, corrected fPINN, and absolute error for Mixed at $t = 1$ with seed 11.

Feature gives a lower mean training loss than Plain and a higher mean test error than Corrected. Lifted reduces the early-time error relative to Plain. The analytic history correction lowers the mean early-time error further, from 2.77×10^{-4} to 1.60×10^{-4} for Mixed and from 2.26×10^{-4} to 1.42×10^{-4} for Localized. Figure 1 shows individual test errors, and Figure 2 compares the exact and Corrected solutions at $t = 1$ for seed 11.

5.4. *Temporal consistency, resolution, and fractional order*

To measure the history quadrature error for a prescribed function, take

$$w(t) = t^{\alpha} + 0.3t^{2\alpha} + 0.2t^{2}.$$

Figure 3 shows the maximum nodal error in the Caputo derivative on uniform grids with $N = 16, 32, 64, 128, 256$. The rates from the last two levels are 1.298 for $\alpha = 0.7$ and 1.100 for $\alpha = 0.9$, in agreement with $2 - \alpha$ in Eq (4.1). The uncorrected maximum defect approaches a nonzero value, consistent with the first-node defect in Eq (4.2).

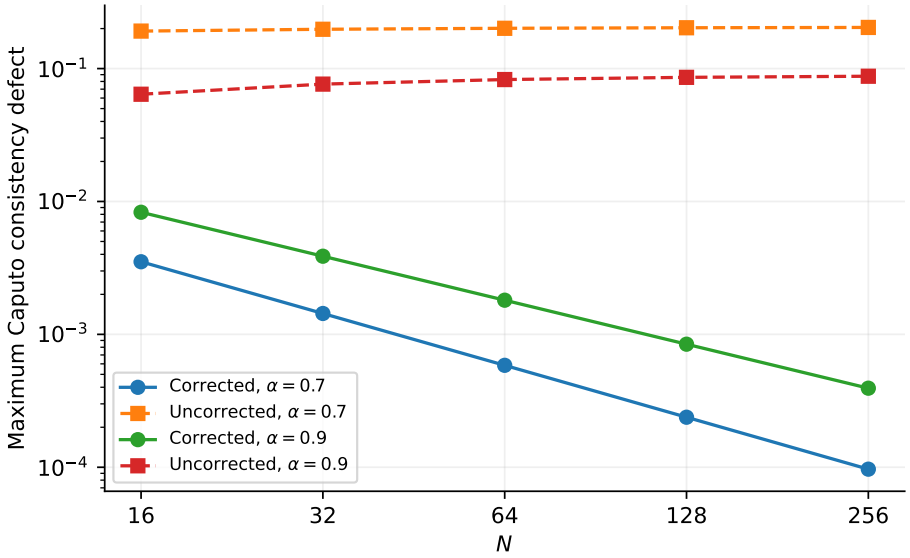


Figure 3. Maximum nodal Caputo-derivative errors for the prescribed history $w(t)$. The corrected formula differentiates both fractional powers analytically.

Table 4. Sensitivity to the time mesh and fractional order in the Mixed test case. All runs use the Corrected method and random seed 11.

α	N	r	E_2	E_∞	$\max_n z_n$
0.7	16	1.000	8.79×10^{-3}	5.89×10^{-3}	5.92×10^{-3}
0.7	32	1.000	4.08×10^{-3}	3.15×10^{-3}	5.32×10^{-3}
0.7	64	1.000	1.86×10^{-3}	1.47×10^{-3}	4.93×10^{-3}
0.7	32	1.857	5.07×10^{-3}	4.30×10^{-3}	4.66×10^{-3}
0.8	32	1.000	5.44×10^{-3}	3.72×10^{-3}	4.95×10^{-3}
0.9	32	1.000	8.23×10^{-3}	5.39×10^{-3}	4.61×10^{-3}

For the time-mesh study, we use $t_n = T(n/N)^r$ with $r = 1$ or $r = (2-\alpha)/\alpha$. The Mixed runs use seed 11 and the same network architecture, optimizer budget, and spatial sample count (Table 4). On uniform meshes, E_2 decreases from 8.79×10^{-3} at $N = 16$ to 1.86×10^{-3} at $N = 64$. These trained-network errors depend on time discretization, optimization, and sampling. In the fractional-order sweep, both the exact solution and the training objective change with α. At $\alpha = 0.7$ and $N = 32$, E_2 increases from 4.08×10^{-3} to 5.07×10^{-3} when the mesh is graded. The larger late-time steps can increase the history error for the regular remainder after the singular powers have been corrected.

Table 5 gives the conventional solver results at $h = 1/64$. The corrected scheme gives a smaller maximum nodal error at each reported N. The rates $\log_2(E_N/E_{2N})$ range from 1.266 to 1.290, close to $2 - \alpha = 1.3$.

Table 5. Conventional solver errors for Mixed at $\alpha = 0.7$ and $h = 1/64$. Errors are maxima over all spatial and positive time nodes.

N	Uncorrected error	Corrected error	Corrected rate
16	1.46×10^{-2}	5.98×10^{-3}	–
32	6.99×10^{-3}	2.47×10^{-3}	1.277
64	3.35×10^{-3}	1.01×10^{-3}	1.290
128	1.60×10^{-3}	4.19×10^{-4}	1.266

5.5. *Verification of the residual bound*

For Corrected in the Mixed test with $N = 32$ and seed 11, Eq (4.9) bounds the network–Newton difference at each time level (Figure 4). At $h = 1/32$, the maximum difference is 1.087×10^{-3} and the maximum paired bound, with both residuals included, is 5.320×10^{-3}. The bound divided by the difference is 4.89. The mean effectivity ratios over the three initializations are 4.64 for Mixed and 2.65 for Localized.

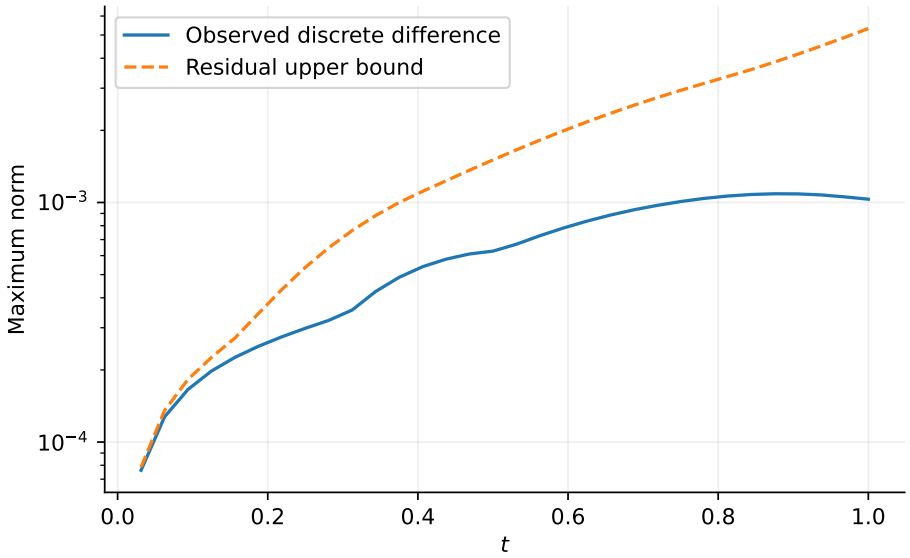


Figure 4. Network–Newton difference $\|W^n - V^n\|_{\infty,h}$ and paired bound from Eq (4.9) for Corrected in the Mixed test with seed 11.

To examine the dependence on the audit grid, we evaluate a single trained network on several grids. For the exact Mixed solution, the constants in Corollary 4.6 satisfy

$$M_t \le 2(0.15) + 6(0.08) = 0.78, \qquad M_4 \le 9.55\pi^4.$$

The coefficient $9.55 = 0.40 + 4.25 + 2.04 + 0.30 + 2.56$ is the sum of the five mode contributions from Eq (5.2) at their maximum time factors on $[0, 1]$.

At $(N_a, K_a) = (32, 32)$, the residual contribution to the PDE nodal bound is 5.320×10^{-3}. The propagated temporal and spatial consistency contributions are 1.2902×10^{-2} and 2.0344×10^{-3}. The three terms sum to the bound 2.0256×10^{-2} in Table 6. The temporal term is the largest and is fixed for this audit grid. When N_a changes, we evaluate the network at every time node of the new history and rebuild the L1 matrix.

Table 6. Residual bounds for a fixed Mixed network trained at $N = 32$ and evaluated on grids with N_a time steps and $K_a = 1/h_a$. The paired bound uses Eq (4.9), and the PDE bound uses Eq (4.10) with the stated analytic constants. Errors and bounds are maxima over interior spatial nodes and positive audit time levels.

N_a	K_a	Discrete difference	Pair bound	Exact nodal error	PDE bound
32	16	1.89×10^{-3}	5.49×10^{-3}	3.15×10^{-3}	2.65×10^{-2}
32	32	1.09×10^{-3}	5.32×10^{-3}	3.19×10^{-3}	2.03×10^{-2}
32	64	1.11×10^{-3}	5.72×10^{-3}	3.20×10^{-3}	1.91×10^{-2}
64	64	2.21×10^{-3}	5.93×10^{-3}	3.20×10^{-3}	1.17×10^{-2}

5.6. Unforced dynamics and certificate-controlled enrichment

For the unforced problem, the discrete bound in Theorem 4.4 supplies the stopping rule for residual-based point enrichment [32, 33]. The third experiment uses $f = 0$, $\alpha = 0.7$, $\varepsilon = 0.1$, $T = 0.5$, and

$$g(x, y) = 0.65 \sin(2\pi x) \sin(\pi y).$$

We compute the lifting coefficients from Eq (2.4), using the expressions in Appendix 6. The smooth, sign-changing initial profile gives a finite-time test of the trajectory and stopping rule. We use the corrected finite-difference solution U_{ref} with $N = 128$ and $h = 1/64$ as the reference. The final relative difference is

$$E^T_{2,\mathrm{ref}} = \frac{\|u_\theta(\cdot, T) - U_{\mathrm{ref}}(T)\|_{\ell^2(\Omega_{1/64})}}{\|U_{\mathrm{ref}}(T)\|_{\ell^2(\Omega_{1/64})}}.$$

For successive coupled refinements $(N, K) = (64, 32), (128, 64), (256, 128)$, the final-time maximum differences on common nodes are 5.885×10^{-4} and 1.481×10^{-4}. These differences give a refinement check for the reference grid.

After the fixed training budget, the loss for seed 33 is 1.264×10^{-5}, compared with 1.386×10^{-5} for seed 11. The final reference difference is larger for seed 33. All three residual bounds exceed the absolute discrete tolerance tol = 0.02.

If $\max_n z_n >$ tol, we evaluate the discrete residual on a candidate grid of spacing $1/24$. We rank the unused locations by $\max_n |\mathcal{R}^n(z)|$ and add the 64 highest-ranked points with their full histories. Training resumes from the current network weights for 150 Adam steps with learning rate 3×10^{-4} and at most

250 L-BFGS iterations. Both enrichment and the continuation control reset the Adam moments and L-BFGS history at the start of each stage. We recompute the bound on the $1/32$ grid and stop when it meets the tolerance or after three stages.

Table 7. Unforced test before and after spatial enrichment. The table gives the final relative difference $E^T_{2,\mathrm{ref}}$ from the $N = 128$, $K = 64$ reference and the maximum discrete error bound $\max_n z_n$ for each seed and stage.

Seed	Stage	Spatial points	Training loss	$E^T_{2,\mathrm{ref}}$	$\max_n z_n$
11	0	128	1.39×10^{-5}	2.50×10^{-3}	2.21×10^{-2}
11	1	192	3.14×10^{-6}	1.00×10^{-3}	8.96×10^{-3}
22	0	128	5.50×10^{-5}	5.90×10^{-2}	7.01×10^{-1}
22	1	192	2.85×10^{-5}	3.26×10^{-3}	2.77×10^{-2}
22	2	256	8.11×10^{-6}	9.54×10^{-4}	8.45×10^{-3}
33	0	128	1.26×10^{-5}	3.59×10^{-2}	2.38×10^{-1}
33	1	192	7.82×10^{-6}	3.06×10^{-3}	2.36×10^{-2}
33	2	256	2.91×10^{-6}	1.23×10^{-3}	8.58×10^{-3}

The discrete bound meets the tolerance after one or two stages for each seed (Table 7 and Figure 5). Final relative reference differences range from 9.536×10^{-4} to 1.226×10^{-3}. On the $1/64$ grid, the reevaluated bounds lie between 8.450×10^{-3} and 9.395×10^{-3}, below 0.02.

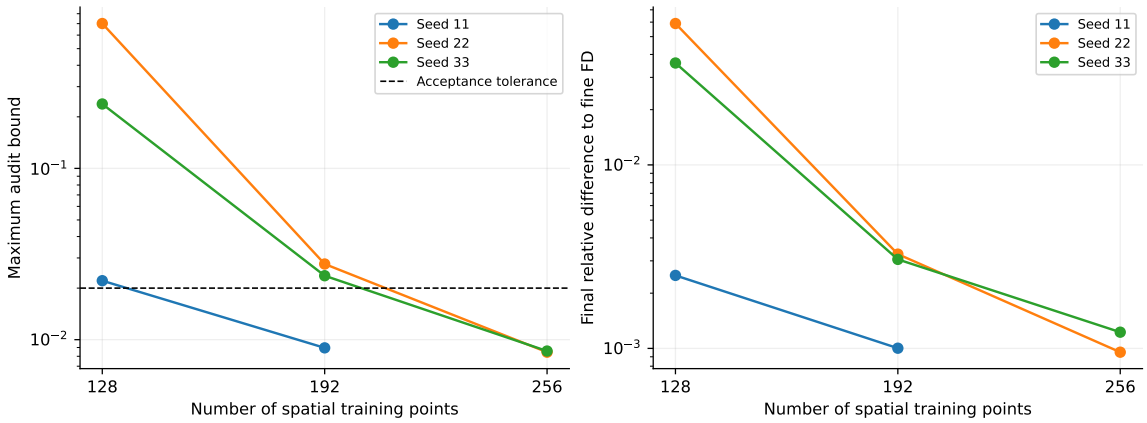


Figure 5. Final reference differences and discrete bounds during spatial enrichment in the unforced test.

Table 8. Continued optimization on the original points and spatial enrichment for the unforced test.

Seed	Phases	Control $E^T_{2,\mathrm{ref}}$	Enriched $E^T_{2,\mathrm{ref}}$	Control bound	Enriched bound
11	1	1.59×10^{-3}	1.00×10^{-3}	1.37×10^{-2}	8.96×10^{-3}
22	2	2.99×10^{-2}	9.54×10^{-4}	3.41×10^{-1}	8.45×10^{-3}
33	2	2.28×10^{-2}	1.23×10^{-3}	2.05×10^{-1}	8.58×10^{-3}

For each seed, both runs resume from the same network for the same number of additional

optimization stages. The control retains the original 128 spatial points. The final reference differences and discrete bounds in Table 8 measure the combined effect of increasing the sample count and selecting the new points by residual magnitude.

6. Conclusions

The initial data and forcing determine the coefficients of t^α and $t^{2\alpha}$ in the trial function. We evaluate the Caputo derivatives of these terms analytically and apply the nonuniform L1 formula to the remainder. Under Eq (2.5), the maximum consistency error of the corrected Caputo approximation over the time nodes is $O(\tau_{\max}^{2-\alpha})$.

The prescribed history correction cancels when two residuals for the same discrete problem are subtracted. Under Eq (4.6), the signs of the L1 coefficients and the one-sided Lipschitz bound for the cubic reaction yield a computable maximum-norm error bound relative to the corrected finite-difference solution. For an inexact reference, the recurrence incorporates the residual norms of both approximations. Including the temporal and spatial consistency defects gives a bound for the error relative to the PDE solution at the grid nodes.

The inverse of the comparison matrix is entrywise nonnegative, so the recurrence can propagate a different consistency bound at each time level. A next step is to derive such bounds on graded meshes when the remainder has limited regularity near $t = 0$. These bounds could guide temporal refinement by showing how the L1 consistency defect contributes to the nodal error bound over time.

Appendix: Lifting for the unforced sine initial condition

For $g = A\sin(2\pi x)\sin(\pi y)$ and $\lambda = 5\pi^2$, let $q = |\nabla g|^2$. Since $\Delta g = -\lambda g$,

$$
\begin{aligned}
s_1 &= \frac{(1-\varepsilon^2\lambda)g - g^3}{\Gamma(1+\alpha)},\\
\Delta s_1 &= \frac{-(1-\varepsilon^2\lambda)\lambda g - 6gq + 3\lambda g^3}{\Gamma(1+\alpha)},\\
s_2 &= \frac{\Gamma(1+\alpha)}{\Gamma(1+2\alpha)}[\varepsilon^2\Delta s_1 + (1-3g^2)s_1].
\end{aligned}
$$

Both coefficients are determined by g and vanish on $\partial\Omega$. We evaluate their remaining spatial derivatives by automatic differentiation.

Author contributions

Lijie Liu: Conceptualization, Methodology, Formal analysis, Writing – review and editing. Leilei Wei: Conceptualization, Methodology, Software, Validation, Writing – original draft.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities of North Minzu University (2026QNPY14).

Conflict of interest

Leilei Wei is a guest editor for Networks and Heterogeneous Media (NHM) and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

1. S. M. Allen, J. W. Cahn, A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening, *Acta Metall.*, **27** (1979), 1085–1095. https://doi.org/10.1016/0001-6160(79)90196-2
2. H. Liu, A. Cheng, H. Wang, J. Zhao, Time-fractional Allen–Cahn and Cahn–Hilliard phase-field models and their numerical investigation, *Comput. Math. Appl.*, **76** (2018), 1876–1892. https://doi.org/10.1016/j.camwa.2018.07.036
3. K. Sakamoto, M. Yamamoto, Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems, *J. Math. Anal. Appl.*, **382** (2011), 426–447. https://doi.org/10.1016/j.jmaa.2011.04.058
4. Q. Du, J. Yang, Z. Zhou, Time-fractional Allen–Cahn equations: Analysis and numerical methods, *J. Sci. Comput.*, **85** (2020), 42. https://doi.org/10.1007/s10915-020-01351-5
5. B. Ji, H. L. Liao, L. Zhang, Simple maximum principle preserving time-stepping methods for time-fractional Allen–Cahn equation, *Adv. Comput. Math.*, **46** (2020), 37. https://doi.org/10.1007/s10444-020-09782-2
6. H. L. Liao, T. Tang, T. Zhou, A second-order and nonuniform time-stepping maximum-principle preserving scheme for time-fractional Allen–Cahn equations, *J. Comput. Phys.*, **414** (2020), 109473. https://doi.org/10.1016/j.jcp.2020.109473
7. T. Tang, H. Yu, T. Zhou, On energy dissipation theory and numerical stability for time-fractional phase-field equations, *SIAM J. Sci. Comput.*, **41** (2019), A3757–A3778. https://doi.org/10.1137/18M1203560
8. H. L. Liao, X. Zhu, J. Wang, The variable-step L1 scheme preserving a compatible energy law for time-fractional Allen–Cahn equation, *Numer. Math. Theory Methods Appl.*, **15** (2022), 1128–1146. https://doi.org/10.4208/nmtma.OA-2022-0011s
9. Z. Mao, Y. Huang, Y. Yu, A. Xiao, H. Yuan, A lattice Boltzmann method for time-fractional Allen–Cahn equation, *Comput. Math. Appl.*, **215** (2026), 122–134. https://doi.org/10.1016/j.camwa.2026.04.030
10. X. Zhang, H. Wang, Z. Luo, L. Wei, A high-accuracy compact finite difference scheme for time-fractional diffusion equations, *Rev. Union Mat. Argent.*, **68** (2025), 589–609. https://doi.org/10.33044/revuma.4665

11. X. Zhang, Y. Feng, L. Wei, High-precision numerical simulation for fractional convection–diffusion–reaction equations, *Math. Methods Appl. Sci.*, **49** (2026), 13928–13945. https://doi.org/10.1002/mma.70756

12. 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

13. Y. Lin, C. Xu, Finite difference/spectral approximations for the time-fractional diffusion equation, *J. Comput. Phys.*, **225** (2007), 1533–1552. https://doi.org/10.1016/j.jcp.2007.02.001

14. B. Jin, R. Lazarov, Z. Zhou, An analysis of the L1 scheme for the subdiffusion equation with nonsmooth data, *IMA J. Numer. Anal.*, **36** (2016), 197–221. https://doi.org/10.1093/imanum/dru063

15. M. Stynes, E. O'Riordan, J. L. Gracia, Error analysis of a finite difference method on graded meshes for a time-fractional diffusion equation, *SIAM J. Numer. Anal.*, **55** (2017), 1057–1079. https://doi.org/10.1137/16M1082329

16. N. Kopteva, Error analysis of the L1 method on graded and uniform meshes for a fractional-derivative problem in two and three dimensions, *Math. Comput.*, **88** (2019), 2135–2155. https://doi.org/10.1090/mcom/3410

17. L. Wei, L. Feng, I. Turner, Z. Mao, F. Liu, Numerical investigation of the 2D unsteady natural convection heat transfer equation with tempered fractional constitutive relationship, *Commun. Nonlinear Sci. Numer. Simul.*, **161** (2026), 110071. https://doi.org/10.1016/j.cnsns.2026.110071

18. H. L. Liao, D. Li, J. Zhang, Sharp error estimate of the nonuniform L1 formula for linear reaction-subdiffusion equations, *SIAM J. Numer. Anal.*, **56** (2018), 1112–1133. https://doi.org/10.1137/17M1131829

19. H. L. Liao, W. McLean, J. Zhang, A discrete Grönwall inequality with applications to numerical schemes for subdiffusion problems, *SIAM J. Numer. Anal.*, **57** (2019), 218–237. https://doi.org/10.1137/16M1175742

20. C. Lubich, Discretized fractional calculus, *SIAM J. Math. Anal.*, **17** (1986), 704–719. https://doi.org/10.1137/0517050

21. M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, **378** (2019), 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

22. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.*, **3** (2021), 422–440. https://doi.org/10.1038/s42254-021-00314-5

23. A. T. Deresse, T. T. Dufera, A deep learning approach: Physics-informed neural networks for solving the 2D nonlinear Sine–Gordon equation, *Results Appl. Math.*, **25** (2025), 100532. https://doi.org/10.1016/j.rinam.2024.100532

24. G. Pang, L. Lu, G. E. Karniadakis, fPINNs: Fractional physics-informed neural networks, *SIAM J. Sci. Comput.*, **41** (2019), A2603–A2626. https://doi.org/10.1137/18M1229845

25. L. Guo, H. Wu, X. Yu, T. Zhou, Monte Carlo fPINNs: Deep learning method for forward and inverse problems involving high dimensional fractional partial differential equations, *Comput. Methods Appl. Mech. Eng.*, **400** (2022), 115523. https://doi.org/10.1016/j.cma.2022.115523

26. X. Kang, Y. Li, Y. Li, J. Hu, K. Zheng, Solving the fractional Allen–Cahn equation and the fractional Cahn–Hilliard equation with the fractional physics-informed neural networks, *Fractal Fract.*, **9** (2025), 773. https://doi.org/10.3390/fractalfract9120773

27. S. Wang, Y. Teng, P. Perdikaris, Understanding and mitigating gradient flow pathologies in physics-informed neural networks, *SIAM J. Sci. Comput.*, **43** (2021), A3055–A3081. https://doi.org/10.1137/20M1318043

28. S. Wang, X. Yu, P. Perdikaris, When and why PINNs fail to train: A neural tangent kernel perspective, *J. Comput. Phys.*, **449** (2022), 110768. https://doi.org/10.1016/j.jcp.2021.110768

29. B. Gao, R. Yao, Y. Li, Physics-informed neural networks with adaptive loss weighting algorithm for solving partial differential equations, *Comput. Math. Appl.*, **181** (2025), 216–227. https://doi.org/10.1016/j.camwa.2025.01.007

30. J. Yu, L. Lu, X. Meng, G. E. Karniadakis, Gradient-enhanced physics-informed neural networks for forward and inverse PDE problems, *Comput. Methods Appl. Mech. Eng.*, **393** (2022), 114823. https://doi.org/10.1016/j.cma.2022.114823

31. S. Wang, S. Sankaran, P. Perdikaris, Respecting causality for training physics-informed neural networks, *Comput. Methods Appl. Mech. Eng.*, **421** (2024), 116813. https://doi.org/10.1016/j.cma.2024.116813

32. L. Lu, X. Meng, Z. Mao, G. E. Karniadakis, DeepXDE: A deep learning library for solving differential equations, *SIAM Rev.*, **63** (2021), 208–228. https://doi.org/10.1137/19M1274067

33. C. Wu, M. Zhu, Q. Tan, Y. Kartha, L. Lu, A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks, *Comput. Methods Appl. Mech. Eng.*, **403** (2023), 115671. https://doi.org/10.1016/j.cma.2022.115671

34. N. Sukumar, A. Srivastava, Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks, *Comput. Methods Appl. Mech. Eng.*, **389** (2022), 114333. https://doi.org/10.1016/j.cma.2021.114333

35. H. Li, Z. Tan, Q. Zou, SC-PIELM: A singularity-capturing physics-informed extreme learning machine for time-fractional PDEs with initial singularities, *Commun. Nonlinear Sci. Numer. Simul.*, **163** (2026), 110789. https://doi.org/10.1016/j.cnsns.2026.110789

36. H. Li, Z. Tan, DSC-PINN: A decomposed singularity-capturing PINN for time-fractional PDEs involving initial singularities, *J. Comput. Phys.*, **561** (2026), 114965. https://doi.org/10.1016/j.jcp.2026.114965

37. S. Kumar, H. K. Dwivedi, Rajeev, Accelerated singularity-aware fractional PINNs for time-fractional phase-field equations: Forward and inverse problems, *J. Comput. Sci.*, **100** (2026), 102969. https://doi.org/10.1016/j.jocs.2026.102969

38. Y. Shin, J. Darbon, G. E. Karniadakis, On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs, *Commun. Comput. Phys.*, **28** (2020), 2042–2074. https://doi.org/10.4208/cicp.OA-2020-0193

39. S. Mishra, R. Molinaro, Estimates on the generalization error of physics-informed neural networks for approximating PDEs, *IMA J. Numer. Anal.*, **43** (2023), 1–43. https://doi.org/10.1093/imanum/drab093

40. T. De Ryck, S. Mishra, Numerical analysis of physics-informed neural networks and related models in physics-informed machine learning, *Acta Numer.*, **33** (2024), 633–713. https://doi.org/10.1017/S0962492923000089

41. N. Kopteva, Pointwise-in-time a posteriori error control for time-fractional parabolic equations, *Appl. Math. Lett.*, **123** (2022), 107515. https://doi.org/10.1016/j.aml.2021.107515

42. S. Franz, N. Kopteva, Pointwise-in-time a posteriori error control for higher-order discretizations of time-fractional parabolic equations, *J. Comput. Appl. Math.*, **427** (2023), 115122. https://doi.org/10.1016/j.cam.2023.115122

43. S. Jiang, J. Zhang, Q. Zhang, Z. Zhang, Fast evaluation of the Caputo fractional derivative and its applications to fractional diffusion equations, *Commun. Comput. Phys.*, **21** (2017), 650–678. https://doi.org/10.4208/cicp.OA-2016-0136



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)