

光纤通信系统中非线性薛定谔方程的复数神经网络求解*

王立斌 张之阳 刘文军†

(北京邮电大学物理科学与技术学院, 北京 100876)

(2026 年 4 月 10 日收到; 2026 年 5 月 20 日收到修改稿)

非线性薛定谔方程用于描述波包在色散介质中传播并受非线性作用影响的核心系统, 在非线性光学、等离子体物理等多个复杂学科中广泛应用. 本文提出一种基于复值神经网络的非线性薛定谔方程求解方法 (PI-CVNN). 该方法能够直接处理复值形式, 天然适应光量子场的复数表示, 并将初始条件、边界条件和方程残差统一纳入物理信息损失函数. PI-CVNN 通过复值权重矩阵显式建模实部与虚部之间的耦合关系, 从而更自然地刻画复包络的演化. 实验结果表明, PI-CVNN 能够有效捕捉光纤孤子传播中的复值动力学行为. 本文研究不仅为光纤通信中的非线性效应建模和优化提供了新的工具, 也拓展了复数神经网络在科学计算领域的应用潜力.

关键词: 非线性薛定谔方程, 光孤子, 物理信息神经网络, 复值神经网络

DOI: 10.7498/aps.75.20260506

CSTR: 32037.14.aps.75.20260506

1 引言

孤子是由非线性效应与色散效应相互作用而形成的稳定波动, 其波形与能量在传播过程中能够保持不变, 因此在流体力学、玻色-爱因斯坦凝聚态 (BECs) 以及等离子体等研究领域占据着重要的研究地位^[1-6]. 在光纤通信领域, Hasegawa 和 Tappert^[7]发现, 当光纤的色散和非线性效应达到平衡时, 光脉冲能够以恒定宽度稳定传播, 从而形成光孤子. 这一发现奠定了高速长距离通信理想载体的理论基础^[8], 其中光孤子的动力学行为可通过非线性薛定谔方程 (nonlinear Schrödinger equation, NLSE) 精确描述.

近年来, 深度学习技术凭借神经网络优异的函数逼近能力, 为光学孤子演化建模提供了新的研究范式. 与有限差分法、有限元方法和分步傅里叶法

等传统数值技术相比, 深度学习方法能够通过数据驱动或物理约束的方式学习复杂系统的潜在动力学规律, 在非线性波方程求解、参数反演和复杂波形重构等问题中展现出良好的应用潜力^[9-14]. 其中, 文献^[15-17]中提出了物理信息神经网络 (physics-informed neural network, PINN), 将控制方程的物理残差嵌入损失函数, 在无需大量标注数据的情况下, 实现了对偏微分方程的高精度求解.

针对非线性波方程及孤子动力学问题, 已有研究对 PINN 及其改进形式进行了较为系统的探索. 文献^[18-22]中将 PINN 和改进 PINN 应用于非线性薛定谔方程数值求解, 成功模拟了孤子、呼吸子、怪波和周期等现象. 此外, 自适应激活函数以及守恒量约束等改进策略也被引入 PINN 框架, 以提升模型的收敛速度、预测精度和物理一致性^[21-24]. 这些研究表明, PINN 类方法在非线性波方程和孤子演化建模中具有良好的适用性^[18-24].

* 国家自然科学基金重点项目 (批准号: 2022YFB4601101)、国家自然科学基金 (批准号: 12261131495) 和北京邮电大学基金 (批准号: 2025JCTP01) 资助的课题.

† 通信作者. E-mail: jungliu@bupt.edu.cn

然而, 现有研究大多基于实值神经网络 (real-valued neural networks, RVNN) 展开. 在求解 NLSE 时, 这类方法通常需要将复数场分解为实部与虚部或幅度与相位分别建模. 此类人为拆分虽然便于计算, 但可能破坏复数场的整体结构, 削弱实部与虚部之间的耦合关系. 尤其对于含有强非线性耦合项的 NLSE, 其动力学演化本质上依赖于复数场的整体表示. 已有研究表明, 复值神经网络在预测精度、收敛速度、优化性能、鲁棒性和泛化能力等方面均具有一定优势^[25,26].

基于此, 本文提出一种基于复值神经网络的物理信息求解框架 (physics-informed complex-valued neural networks, PI-CVNN). 该方法直接在复数空间中对光场进行建模, 并将初始条件、边界条件以及 NLSE 残差统一纳入网络训练过程, 从而获得高精度的全域近似解. 与传统实值神经网络相比, PI-CVNN 无需对复值光场进行实虚分解, 能够更自然地刻画光场的相位特征及其几何结构.

本文的结构如下: 第 2 节介绍 PINN 的基本求解框架及 PI-CVNN 的模型构建方法; 第 3 节给出用于描述光纤中超短脉冲传播动力学的非线性薛定谔方程及其孤子解; 第 4 节使用 PI-CVNN 和传统 PINN 方法对非线性薛定谔方程的正问题进行求解; 第 5 节总结实验结果并给出未来展望.

2 PI-CVNN 方法

2.1 问题定义

本文面向非线性薛定谔方程所描述的复值动力学系统, 构建一种物理信息复值神经网络求解框架. 设时间变量为 t , 传播距离变量为 x , 计算区域记为 $\Omega = \mathcal{T} \times \mathcal{X}$, 待求复包络为 $U(t, x) \in \mathbb{C}$. 为了保持方法框架的一般性, 首先考虑如下复值偏微分方程约束:

$$\mathcal{N}[U; \boldsymbol{\mu}](t, x) = 0, \quad (t, x) \in \Omega, \quad (1)$$

其中, $\mathcal{N}[\cdot]$ 表示复值微分算子, $\boldsymbol{\mu}$ 表示方程参数. 初始条件记为 $U(t, x_0) = U_0(t)$, 边界条件记为 $\mathcal{B}[U](t, x) = g_b(t, x)$. 基于神经网络的求解方法, 求解目标是在给定初始边界约束的条件下, 获得满足物理方程的近似数值解. 对于正问题, $\boldsymbol{\mu}$ 为已知参数; 对于反问题, $\boldsymbol{\mu}$ 可作为待估计变量与神经网络参数共同优化. 本文关注的是正问题的求解, 具体是描述

光纤中脉冲传输的非线性薛定谔方程的正向求解.

2.2 复值神经网络方法

为提高自动微分中高阶导数计算的数值稳定性, 本文首先将原始物理坐标线性归一化至 $[-1, 1]^2$, 如下所示:

$$\tilde{t} = 2 \frac{t - t_{\min}}{t_{\max} - t_{\min}} - 1, \quad \tilde{x} = 2 \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1. \quad (2)$$

记归一化坐标向量为 $\tilde{\mathbf{s}} = (\tilde{t}, \tilde{x})^\top \in \mathbb{R}^2$.

PI-CVNN 使用参数为 θ 的复值神经网络 $U_\theta(t, x)$ 逼近真实复值解 $U(t, x)$, 即

$$U(t, x) \approx U_\theta(t, x) = u_\theta(t, x) + i v_\theta(t, x), \quad (3)$$

其中, 参数 θ 为复数, $u_\theta(t, x)$ 和 $v_\theta(t, x)$ 分别表示网络输出的实部与虚部, i 为虚数单位. 由于输入坐标为实值变量, 而网络采用复值特征传播, 本文使用复值线性输入层将归一化坐标嵌入复值特征空间:

$$h^0 = W_e \tilde{\mathbf{s}} + b_e \in \mathbb{C}^{n_0}, \quad (4)$$

其中 $W_e \in \mathbb{C}^{n_0 \times 2}$, $b_e \in \mathbb{C}^{n_0}$, n_0 为输入嵌入维度. 设 PI-CVNN 包含 L 个复值线性层, 其中前 $L-1$ 层为隐藏层, 最后一层为线性输出层. 隐藏层传播过程为

$$h^l = \sigma(W^l h^{l-1} + b^l), \quad l = 1, 2, \dots, L-1, \quad (5)$$

其中 $W^l \in \mathbb{C}^{n_l \times n_{l-1}}$, $b^l \in \mathbb{C}^{n_l}$, $\sigma(\cdot)$ 为复值激活函数. 输出层定义为

$$\hat{U}_\theta(\tilde{t}, \tilde{x}) = W^L h^{L-1} + b^L \in \mathbb{C}, \quad (6)$$

于是有

$$U_\theta(t, x) = \hat{U}_\theta(\tilde{t}, \tilde{x}). \quad (7)$$

为说明复值神经网络的结构, 令

$$W^l = W_r^l + i W_i^l, \quad h^{l-1} = h_r^{l-1} + i h_i^{l-1}. \quad (8)$$

则有

$$\begin{aligned} W^l h^{l-1} &= (W_r^l h_r^{l-1} - W_i^l h_i^{l-1}) \\ &\quad + i (W_r^l h_i^{l-1} + W_i^l h_r^{l-1}), \end{aligned} \quad (9)$$

等价的矩阵表示形式为

$$\begin{bmatrix} \text{Re}(W^l h^{l-1}) \\ \text{Im}(W^l h^{l-1}) \end{bmatrix} = \begin{bmatrix} W_r^l & -W_i^l \\ W_i^l & W_r^l \end{bmatrix} \begin{bmatrix} h_r^{l-1} \\ h_i^{l-1} \end{bmatrix}. \quad (10)$$

该块矩阵结构表明, 复值线性层通过与复数乘法一致的块结构, 在网络内部显式引入实部与虚部的交叉作用, 从而为复值解的建模提供结构化归纳偏置.

2.3 激活函数和复值残差

本文研究的是从实变量坐标 (t, x) 到复值物理场 $U(t, x)$ 的映射, 而非复变函数意义下的全纯映射. 因此, 复值激活函数不要求满足全纯性. 实际实现中, 可采用分离型激活函数:

$$\sigma(a + ib) = \phi(a) + i\phi(b), \quad (11)$$

其中, $\phi(\cdot)$ 为实值非线性激活函数. 考虑到 NLSE 物理残差中包含高阶导数, 激活函数应具有较好的平滑性. 本文采用分离型双曲正切函数:

$$\sigma(a + ib) = \tanh(a) + i \tanh(b). \quad (12)$$

物理信息神经网络方法的关键在于将偏微分方程约束嵌入神经网络训练目标. 内嵌物理信息的复值神经网络也是如此. 将神经网络复值近似解 $U_\theta(t, x)$ 代入物理算子, 可得到复值残差

$$R_\theta(t, x; \boldsymbol{\mu}) = \mathcal{N}[U_\theta; \boldsymbol{\mu}](t, x). \quad (13)$$

以经典 NLSE 为例,

$$i \frac{\partial U}{\partial x} + \frac{\beta}{2} \frac{\partial^2 U}{\partial t^2} + \gamma |U|^2 U = 0, \quad (14)$$

其残差可写为

$$R_\theta(t, x; \boldsymbol{\mu}) = i \frac{\partial U_\theta}{\partial x} + \frac{\beta}{2} \frac{\partial^2 U_\theta}{\partial t^2} + \gamma |U_\theta|^2 U_\theta. \quad (15)$$

经典 NLSE 网络近似解 $U_\theta(t, x) = u_\theta(t, x) + iv_\theta(t, x)$, 其关于实变量的偏导数可由自动微分分别作用于 u_θ 和 v_θ 的计算图, 并组合为复值导数得到. 例如,

$$\frac{\partial U_\theta}{\partial x} = \frac{\partial u_\theta}{\partial x} + i \frac{\partial v_\theta}{\partial x}, \quad \frac{\partial^2 U_\theta}{\partial t^2} = \frac{\partial^2 u_\theta}{\partial t^2} + i \frac{\partial^2 v_\theta}{\partial t^2}. \quad (16)$$

考虑到网络输入的归一化, 残差中的导数需要通过链式法则恢复到原始坐标尺度. 即有

$$\frac{\partial U_\theta}{\partial x} = \frac{2}{x_{\max} - x_{\min}} \frac{\partial \hat{U}_\theta}{\partial \tilde{x}}, \quad (17)$$

以及

$$\frac{\partial^k U_\theta}{\partial t^k} = \left(\frac{2}{t_{\max} - t_{\min}} \right)^k \frac{\partial^k \hat{U}_\theta}{\partial \tilde{t}^k}, \quad k = 1, 2, 3. \quad (18)$$

本文在构造物理残差时均使用链式法则修正后的导数, 以保证归一化处理不改变原始方程的物理含义.

2.4 损失函数设计

U_θ , R_θ 以及观测数据均为复值, 本文采用复数模平方度量误差. 设定域内部配置点、初始条

件点、边界条件点和观测数据点分别为 $\mathcal{D}_f = \{(t_f^j, x_f^j)\}_{j=1}^{N_f}$, $\mathcal{D}_{ic} = \{(t_{ic}^j)\}_{j=1}^{N_{ic}}$, $\mathcal{D}_{bc} = \{(t_{bc}^j, x_{bc}^j)\}_{j=1}^{N_{bc}}$, $\mathcal{D}_d = \{(t_d^j, x_d^j, U_d^j)\}_{j=1}^{N_d}$. 方程残差损失可写为

$$\mathcal{L}_f = \frac{1}{N_f} \sum_{j=1}^{N_f} |R_\theta(t_f^j, x_f^j; \boldsymbol{\mu})|^2. \quad (19)$$

初始条件损失为

$$\mathcal{L}_{ic} = \frac{1}{N_{ic}} \sum_{j=1}^{N_{ic}} |U_\theta(t_{ic}^j, x_0) - U_0(t_{ic}^j)|^2. \quad (20)$$

边界条件损失表示为

$$\mathcal{L}_{bc} = \frac{1}{N_{bc}} \sum_{j=1}^{N_{bc}} |\mathcal{B}[U_\theta](t_{bc}^j, x_{bc}^j) - g_b(t_{bc}^j, x_{bc}^j)|^2. \quad (21)$$

若引入观测数据, 则数据损失为

$$\mathcal{L}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} |U_\theta(t_d^j, x_d^j) - U_d^j|^2. \quad (22)$$

对微分方程的求解转化为对网络参数的优化问题:

$$\min_{\theta, \boldsymbol{\mu}} \mathcal{L}(\theta, \boldsymbol{\mu}) = \omega_f \mathcal{L}_f + \omega_{ic} \mathcal{L}_{ic} + \omega_{bc} \mathcal{L}_{bc} + \omega_d \mathcal{L}_d, \quad (23)$$

其中, ω_f , ω_{ic} , ω_{bc} 和 ω_d 为损失权重. 对于正问题, 方程参数 $\boldsymbol{\mu}$ 已知, 仅优化网络参数 θ .

需要指出的是, 本文聚焦于物理信息复值特征网络的构造和正问题求解, 为避免自适应权重机制引入额外变量, 在实验中各损失项均不考虑加权情况. 同时在正问题求解时不引入观测数据. 即在正问题求解时, 加权系数 $\omega_f = \omega_{ic} = \omega_{bc} = 1$, $\omega_d = 0$.

2.5 参数初始化方法

本文对传统 PINN 与 PI-CVNN 采用方差尺度可比的初始化策略. 对于传统双输出实值 PINN, 实值权重采用 Xavier 均匀初始化, 偏置初始化为零. 若第 l 层权重的输入维度和输出维度分别为 n_{l-1} 和 n_l , 则

$$W_{ij}^l \sim \mathcal{U}(-a_l, a_l), \quad a_l = \sqrt{6/(n_{l-1} + n_l)}. \quad (24)$$

对于 PI-CVNN 的复值权重 $W^l = W_r^l + iW_i^l$ 来说, 为使复值线性层的整体方差尺度与实值 Xavier 初始化保持可比, 本文分别对实部和虚部采用缩放 Xavier 初始化:

$$W_{r,ij}^l \sim \mathcal{U}\left(-\frac{a_l}{\sqrt{2}}, \frac{a_l}{\sqrt{2}}\right), \quad W_{i,ij}^l \sim \mathcal{U}\left(-\frac{a_l}{\sqrt{2}}, \frac{a_l}{\sqrt{2}}\right). \quad (25)$$

复值偏置的实部和虚部均初始化为零.

3 光纤中超短脉冲传播模型

超短 (飞秒) 脉冲的传播具有重要研究意义, 被广泛应用于超高比特率光通信、超快物理过程和光谱学等领域. 飞秒脉冲在光纤中的传播可通过高阶非线性薛定谔方程描述, 该方程包含二阶色散、自相位调制、三阶色散、自陡峭和受激拉曼散射等效应^[27]. 飞秒脉冲在均匀光纤中的传播可以通过以下归一化高阶非线性薛定谔方程描述:

$$q_x = i\alpha_2 q_{tt} + i\gamma_1 |q|^2 q + \alpha_3 q_{ttt} + \gamma_2 (|q|^2 q)_t + \gamma_3 q (|q|^2)_t, \quad (26)$$

其中 $q(t, x)$ 是电场的慢变复包络, x 和 t 分别表示归一化传播距离和时间. 实参数 α_2 , γ_1 , α_3 , γ_2 , γ_3 分别控制群速度色散、克尔非线性 (自相位调制)、三阶色散、自陡峭效应和受激拉曼散射效应的强度.

引入振幅标度变化 $u(t, x) = \mu q(t, x)$ 和选择特殊参数 $\gamma_1 = 2\mu^2 \alpha_2$, $\gamma_2 = 6\mu^2 \alpha_3$, $\gamma_3 = -\gamma_2$, 方程 (26) 化为如下著名 Hirota 可积模型^[28]:

$$iu_x + \alpha_2 (u_{tt} + 2|u|^2 u) - i\alpha_3 (u_{ttt} + 6|u|^2 u_t) = 0, \quad (27)$$

其中, α_2 和 α_3 为常数参数, 分别控制二阶色散和三阶色散相关高阶效应. 当 $\alpha_3 = 0$ 时, 模型退化为只包含二阶色散和克尔非线性的标准 NLSE; 当 $\alpha_3 \neq 0$ 时, 模型描述常系数高阶 NLSE 中的脉冲传播行为.

本文研究的是 NLSE 中单孤子的演化问题. 为了衡量 PI-CVNN 的求解精度、构造初始条件和 Dirichet 型边界条件, 本文利用 Hirota 双线性方法得到 (27) 式的单孤子解析解. 其形式为

$$u(t, x) = \frac{e^\tau}{1 + \frac{e^{\tau + \tau^*}}{(\eta + \eta^*)^2}}, \quad (28)$$

其中 η 是任意复数参数, τ^* 表示 τ 的复共轭, ϕ_0 是实常数, τ 定义为

$$\tau = \int (\alpha_2 \eta^2 + \alpha_3 \eta^3) dx + \eta t + \phi_0. \quad (29)$$

实验中, 取复参数 $\eta = 0.8 - i$, 常实数 $\phi_0 = 2$. 基于不同色散系数 α_2 , α_3 , 本文考虑以下两类问题求解场景.

第一种情形为 $\alpha_2 = 0.5$, $\alpha_3 = 0$, 此时方程退

化为仅包含二阶色散和克尔非线性效应的非线性薛定谔方程. 由 (28) 式和 (29) 式可得单孤子解析解为

$$u(t, x) = \frac{2.56e^{0.8t+0.8x+2-(0.18x+t)i}}{2.56 + e^{1.6t+1.6x+4}}, \quad (30)$$

设置求解区域为 $[t, x] \in [-8, 3] \times [0, 1]$.

第二种情形为 $\alpha_2 = 1$, $\alpha_3 = 0.6$, 此时方程为包含高阶色散的非线性薛定谔方程. 由 (28) 式和 (29) 式可得对应单孤子解析解为

$$u(t, x) = \frac{2.56e^{0.8t+0.4672x+2-(0.912x+t)i}}{2.56 + e^{1.6t+0.9344x+4}}, \quad (31)$$

设置求解区域为 $[t, x] \in [-10, 5] \times [0, 2]$.

3.1 PICVNN 损失函数

根据 PI-CVNN 方法框架, 使用参数为 θ 的复值神经网络 $U_\theta(t, x)$ 逼近方程复值解 $u(t, x)$:

$$u(t, x) \approx U_\theta(t, x) = u_\theta(t, x) + iv_\theta(t, x), \quad (32)$$

其中, $u(t, x)$ 表示方程复数解, $u_\theta(t, x)$ 和 $v_\theta(t, x)$ 分别表示网络输出的实部与虚部, i 为虚数单位. 将 U_θ 代入方程 (27), 得到复值残差表示为

$$R_\theta(t, x) = iU_{\theta,x} + \alpha_2 (U_{\theta,tt} + 2|U_\theta|^2 U_\theta) - i\alpha_3 (U_{\theta,ttt} + 6|U_\theta|^2 U_{\theta,t}), \quad (33)$$

式中 $U_{\theta,x}$, $U_{\theta,tt}$, $U_{\theta,ttt}$ 均表示经链式法则恢复到原始坐标尺度后的导数. 其中

方程残差损失为

$$\mathcal{L}_f^{\text{PI-CVNN}} = \frac{1}{N_f} \sum_{j=1}^{N_f} |R_\theta(t_f^j, x_f^j)|^2. \quad (34)$$

初始条件损失为

$$\mathcal{L}_{ic} = \frac{1}{N_{ic}} \sum_{j=1}^{N_{ic}} |U_\theta(t_{ic}^j, x_0) - u_{\text{exact}}(t_{ic}^j, x_0)|^2. \quad (35)$$

边界条件损失为

$$\mathcal{L}_{bc} = \frac{1}{N_{bc}} \sum_{j=1}^{N_{bc}} |U_\theta(t_{bc}^j, x_{bc}^j) - u_{\text{exact}}(t_{bc}^j, x_{bc}^j)|^2. \quad (36)$$

若引入观测数据, 则数据损失为

$$\mathcal{L}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} |U_\theta(t_d^j, x_d^j) - u_{\text{exact}}(t_d^j, x_d^j)|^2. \quad (37)$$

本文聚焦于物理信息复值特征网络的构造, 在实验中各损失项均不考虑加权情况, 同时在正问题求解时未使用配置点观测数据. 总损失可统一表示为

$$\mathcal{L}^{\text{PI-CVNN}} = \mathcal{L}_f^{\text{PI-CVNN}} + \mathcal{L}_{\text{ic}} + \mathcal{L}_{\text{bc}}. \quad (38)$$

3.2 与实值神经网络的联系

对于传统实值 PINN, 其实部和虚部的组合输出仍记为 $U_\theta(t, x) = u_\theta(t, x) + i v_\theta(t, x)$, 并定义 $\rho_\theta(t, x) = u_\theta^2(t, x) + v_\theta^2(t, x)$, 则 (33) 式实部虚部分别为

$$\begin{aligned} f_r &= \text{Re}(R_\theta) = -v_{\theta,x} + \alpha_2(u_{\theta,tt} + 2\rho_\theta u_\theta) \\ &\quad + \alpha_3(v_{\theta,ttt} + 6\rho_\theta v_{\theta,t}), \\ f_i &= \text{Im}(R_\theta) = u_{\theta,x} + \alpha_2(v_{\theta,tt} + 2\rho_\theta v_\theta) \\ &\quad - \alpha_3(u_{\theta,ttt} + 6\rho_\theta u_{\theta,t}). \end{aligned} \quad (39)$$

传统 PINN 的残差损失写为

$$\mathcal{L}_f^{\text{PINN}} = \frac{1}{N_f} \sum_{j=1}^{N_f} \left[\lambda_1 \left(f_r(t_f^j, x_f^j) \right)^2 + \lambda_2 \left(f_i(t_f^j, x_f^j) \right)^2 \right]. \quad (40)$$

当 $\lambda_1 = \lambda_2 = 1$ 时, 有

$$\mathcal{L}_f^{\text{PINN}} = \frac{1}{N_f} \sum_{j=1}^{N_f} |R_\theta(t_f^j, x_f^j)|^2. \quad (41)$$

同时初始条件、边界条件损失、可能的观测数据在复数模平方意义下与 PI-CVNN 的对应项保持一致, 因而传统 PINN 的总损失可统一表示为

$$\mathcal{L}^{\text{PINN}} = \mathcal{L}_f^{\text{PINN}} + \mathcal{L}_{\text{ic}} + \mathcal{L}_{\text{bc}} + \mathcal{L}_d. \quad (42)$$

由此可见, 在采用复数模平方误差且实部、虚部等权处理的情况下, PI-CVNN 与传统双输出实值 PINN 在损失函数形式上保持一致. 本文关注的重点并非损失函数设计差异, 而是复值参数化结构本身的作用. 传统 PINN 以普通实值网络同时预测实部和虚部, 而 PI-CVNN 通过复值权重矩阵及其对应的块矩阵结构, 在前向传播中显式引入实部与虚部的交叉耦合, 从而为复包络的幅值-相位联合演化提供结构化归纳偏置. 换言之, PI-CVNN 的潜在优势主要体现在参数组织方式、表达效率上, 而不应简单理解为损失函数层面的优势.

4 实验

4.1 实验设置

实验围绕所提出的内嵌物理信息复值神经网络 PI-CVNN 展开, 目标是验证其在光纤脉冲两类传输模型正问题求解中的可行性与有效性.

为评价 PI-CVNN 的求解效果, 首先通过双线性法构造精确解来验证. 同时设置传统双输出实值 PINN 作为方法对比基线. 考虑到复值网络中的一个复权重对应实部和虚部两个实参数, 在相同网络宽度下 PI-CVNN 通常具有更高参数量. 本文进一步构造参数量近似对齐的 PINN-param, 即通过增大实值 PINN 的隐藏层宽度, 使其参数量与 PI-CVNN 处于相近水平. 客观地评价 PI-CVNN 在复值脉冲传播建模中的作用.

本文的主要目标是考察复值参数化结构本身对 NLSE 正问题求解的影响, 为避免自适应权重机制引入额外变量, 本文中的三类模型均不使用损失加权策略. 同时, 三类模型在训练点、初边值点、测试点、采样策略、优化策略和训练轮数也保持一致.

所有模型输入坐标均归一化至 $[-1, 1]^2$, 网络均采用全连接结构, 基激活函数为 $\tanh$, 配置点采用拉丁超立方采样. 参与物理残差约束的配置点数设置为 $N_f = 5000$, 初始条件点数为 $N_{\text{ic}} = 50$, 边界条件点数为 $N_{\text{bc}} = 100$. 初始条件、边界条件和测试标签均由解析解 $u_{\text{exact}}(t, x)$ 给出. 测试阶段在 256×256 的均匀网格上评估模型精度. 训练过程采用 Adam 与 L-BFGS 结合的混合优化策略. Adam 优化器学习率设为 10^{-3} , 迭代 1000 步; L-BFGS 优化器学习率设为 1.0, 最大迭代次数设为 3000, 历史更新步数设为 50, 收敛容差设为 10^{-6} . 所有实验均进行 5 次重复实验, 随机种子为 $\{0, 1, 2, 3, 4\}$. 所有实验均采用 float32 精度, 并在单张 Nvidia RTX 3060 12 GB GPU 上完成.

在具体实现中, PI-CVNN 采用分离型复值激活函数

$$\sigma(a + ib) = \tanh(a) + i \tanh(b), \quad (43)$$

即分别对复值特征的实部和虚部施加 $\tanh$ 非线性. 复值权重 $W = W_r + iW_i$ 的实部和虚部分别采用缩放 Xavier 初始化:

$$W_r, W_i \sim \mathcal{U}\left(-\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right), \quad (44)$$

以保持与实值 Xavier 初始化可比的方差尺度. 复值偏置的实部和虚部均初始化为零.

在 PyTorch 实现中, PI-CVNN 的复值参数以实部和虚部两个可训练张量形式存储, 即 $W = W_r + iW_i$. 前向传播时按照复数乘法规则构造复值线性变换, 并输出复值近似解 $U_\theta = u_\theta + i v_\theta$. 由于输入变量 t, x 为实值坐标, 物理残差中的各阶导数

通过 PyTorch 自动微分分别作用于 u_θ 和 v_θ , 再组合为复值导数. 损失函数采用复数模平方形式, 最终为实值标量, 因此可以通过标准反向传播更新 W_r 和 W_i . 本文未使用额外的复数梯度截断机制.

4.2 评价指标

本文采用复值相对 L_2 误差评价整体预测精度:

$$E_{L_2} = \frac{\|U_\theta - u_{\text{exact}}\|_2}{\|u_{\text{exact}}\|_2}. \quad (45)$$

为进一步考察模型对孤子幅值结构的刻画能力, 本文报告幅值误差:

$$E_{\text{amp}} = \frac{\| |U_\theta| - |u_{\text{exact}}| \|_2}{\| |u_{\text{exact}}| \|_2}. \quad (46)$$

考虑到相位存在分支跳变, 本文采用复共轭主值相位差:

$$\Delta\phi_j = \text{angle} \left(U_\theta(t_j, x_j) \overline{u_{\text{exact}}(t_j, x_j)} \right), \quad (47)$$

并定义相位误差为

$$E_{\text{phase}} = \left(\frac{1}{|\mathcal{S}_{\text{test}}|} \sum_{(t_j, x_j) \in \mathcal{S}_{\text{test}}} |\Delta\phi_j|^2 \right)^{1/2}. \quad (48)$$

4.3 参数退化下求解

本节考虑参数退化情形: $\alpha_2 = 0.5$, $\alpha_3 = 0$. 此时方程退化为标准 NLSE. 该实验用于验证 PI-CVNN 在基础复值孤子传播问题中的求解能力. 本节采用 5 层每层为 40 个神经元的复值神经网络求解. 实验详细参数设置见表 A1. 选取 5 组随机种子进行重复实验, 三类模型的平均误差结果如表 1 所列. 详细统计结果见表 A2. 当随机种子为 0, 幅值和相位的预测结果如图 1 和图 2 所示. 图 1 展示了幅值精确值、幅值预测值及幅值绝对误差分布热力图和三个传播位置上的幅值快照. 图 2 展示相位精确值、相位预测值及复共轭主值相位绝对误差分布图.

表 1 参数退化情形下的 PI-CVNN 求解误差

Table 1. Forward solving errors in the degenerate-parameter case.

模型	复值解 L_2 误差	振幅误差	相位误差	最终损失	训练时间/s
PINN	1.421×10^{-3}	1.023×10^{-3}	6.738×10^{-3}	1.923×10^{-7}	213
PINN-param	1.254×10^{-3}	9.273×10^{-4}	6.356×10^{-3}	1.152×10^{-7}	329
PI-CVNN	1.278×10^{-3}	9.160×10^{-4}	6.869×10^{-3}	1.035×10^{-7}	814

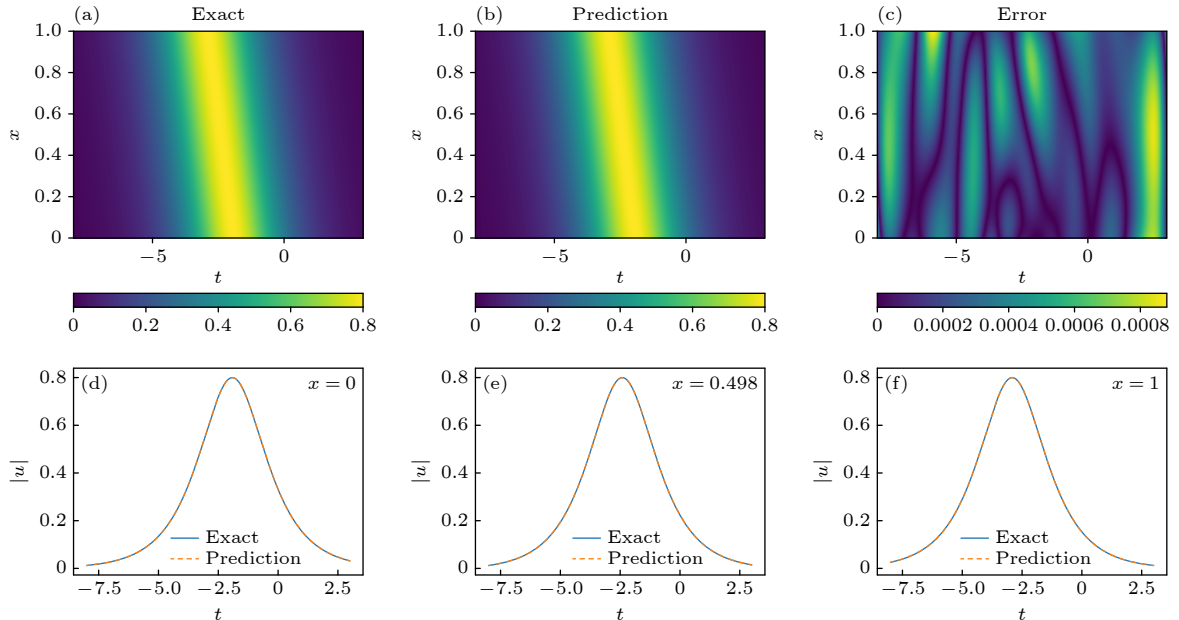


图 1 PI-CVNN 求解退化 NLSE 的幅值结果 (a) 幅值精确解热力图; (b) 幅值预测解热力图; (c) 幅值误差分布热力图; (d) 不同传播距离处的幅值快照

Fig. 1. Amplitude results of PI-CVNN for the degenerate NLSE: (a) Heatmap of exact amplitude; (b) heatmap of predicted amplitude; (c) heatmap of amplitude error distribution; (d) amplitude snapshots at different propagation distances.

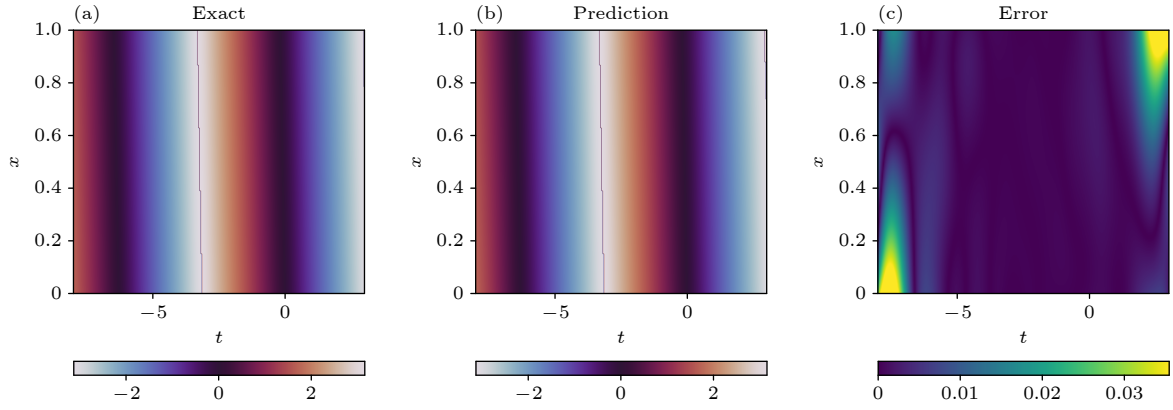


图 2 PI-CVNN 求解退化 NLSE 的相位结果 (a) 相位精确解热力图; (b) 相位预测解热力图; (c) 相位误差分布热力图

Fig. 2. Phase results of PI-CVNN for the degenerate NLSE: (a) Heatmap of exact phase; (b) heatmap of predicted phase; (c) heatmap of phase error distribution.

表 2 常系数情况下 PI-CVNN 的求解误差
Table 2. Forward solving errors in the constant-coefficient case.

模型	复值解 L_2 误差	振幅误差	相位误差	总损失	训练时间/s
PINN	2.262×10^{-3}	1.562×10^{-3}	1.847×10^{-2}	8.488×10^{-8}	251
PINN-param	3.080×10^{-3}	2.237×10^{-3}	3.395×10^{-2}	2.735×10^{-7}	327
PI-CVNN	2.136×10^{-3}	1.674×10^{-3}	2.326×10^{-2}	7.597×10^{-8}	1134

可以看出, 在参数退化的场景下, PI-CVNN 模型能较好逼近解析孤子解, 整体误差均处于 10^{-3} 量级. 与相同结构的实值神经网络有着同水平的拟合效果. 从上述理论分析可知, 复值线性层引入的实虚部耦合计算显然会带来更高的参数计算开销. PI-CVNN 训练时间较长, 这是符合预期的.

4.4 常系数下求解

本节考虑常系数高阶情形: $\alpha_2 = 1$, $\alpha_3 = 0.6$. 该实验用于检验 PI-CVNN 在高阶 NLSE 场景下的正问题求解能力. 本节采用 5 层每层为 80 个神经元的复值神经网络进行求解. 实验详细参数设置见表 A3. 选取 5 组随机种子进行重复实验, 三类模型的平均误差结果如表 2 所列. 详细统计结果见表 A4. 当随机种子为 0, 幅值和相位的预测结果如图 3 和图 4 所示. 图 3 展示了幅值精确值、幅值预测值及幅值绝对误差分布热力图和三个传播位置上的幅值快照. 图 4 展示相位精确值、相位预测值及复共轭主值相位绝对误差分布图.

从表 2 可以看出, 在常系数高阶 NLSE 场景下, PI-CVNN 取得了最低的相对 L_2 误差, 说明其在整体复值解逼近和物理残差约束方面具有一定优势. 与参数量近似对齐的 PINN-param 相比, PI-CVNN 在所有主要误差指标和最终损失上均表

现更好, 表明仅增加实值网络宽度并不必然提升高阶 NLSE 的求解效果, 复值结构中的实虚部耦合可能对复杂复值场建模具有积极作用. 从计算代价看, PI-CVNN 的平均训练时间显著高于两个实值基线. 该结果也表明, PI-CVNN 的复值结构会带来更高训练成本, 符合预期.

5 结 论

本文提出一种面向 NLSE 复值动力学系统的 PI-CVNN 求解框架, 直接在复值特征空间中表示光场包络, 通过复值线性层建模实部与虚部之间的耦合关系. 为光纤通信中的非线性效应建模和优化提供了新的工具. 将 PI-CVNN 具体应用于 Hirota 型高阶 NLSE 的正问题求解, 基于解析孤子解构造初始条件、边界条件和测试误差评估, 从而在可控条件下验证其求解能力. 通过与传统双输出 PINN 及参数量近似对齐 PINN 进行比较, 分析复值结构对整体误差、幅值误差、相位误差、物理残差收敛和训练时间的影响, 指出 PI-CVNN 在复值结构表达上的潜力及其计算成本.

在实验中, PI-CVNN 的复值 L_2 误差、幅度误差和相位误差均保持在 10^{-3} 量级, 说明复值神经网络结合物理残差约束后, 能够较好逼近光纤脉冲

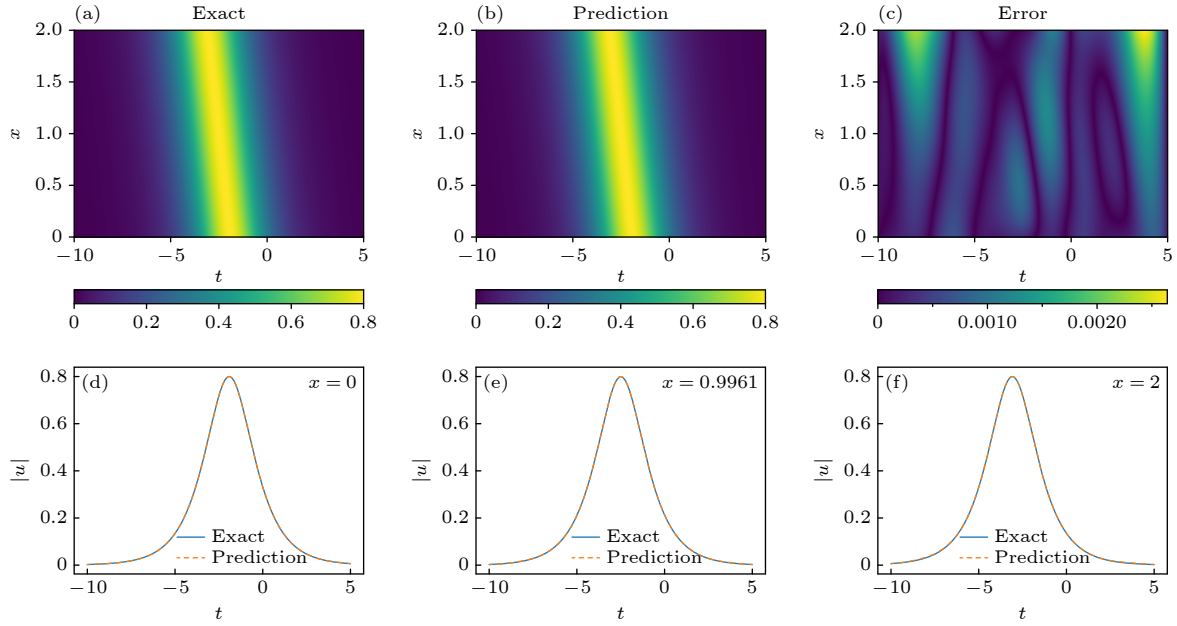


图 3 PI-CVNN 求解常系数 NLSE 的幅值结果 (a) 幅值预测图; (b) 幅值误差分布图; (c) 不同传播距离处的幅值快照

Fig. 3. Amplitude results of PI-CVNN for the constant-coefficient NLSE: (a) Amplitude prediction; (b) amplitude error distribution; (c) amplitude snapshots at different propagation distances.

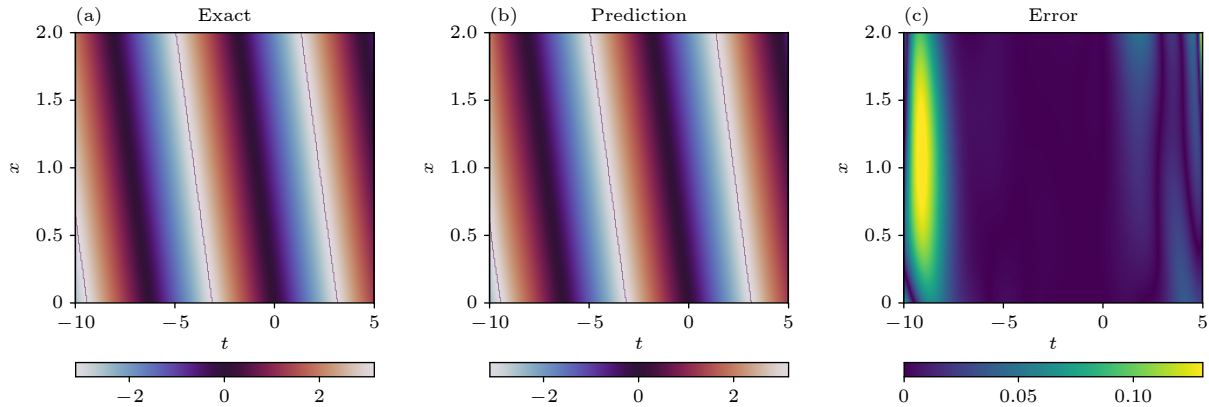


图 4 PI-CVNN 求解常系数 NLSE 的相位结果 (a) 相位精确解热力图; (b) 相位预测解热力图; (c) 相位误差分布热力图

Fig. 4. Phase results of PI-CVNN for the constant-coefficient NLSE: (a) Heatmap of exact phase; (b) heatmap of predicted phase; (c) heatmap of phase error distribution.

传播中的复值孤子解. 在常系数高阶 NLSE 情形下, PI-CVNN 相比参数量近似对齐的 PINN 表现出更好的整体复值逼近能力和物理残差收敛效果. 说明复值线性层所引入的实虚部耦合结构有助于复杂复值场的整体建模. 需要指出的是, 复值结构在增强表达能力的同时, 不可避免也带来了更高计算成本.

本文对 PI-CVNN 的研究仍处于初步阶段, 诸多问题亟待深入探讨. 其一, 本文聚焦于复值神经网络构建, 对损失进行未加权处理, 而实际中未加权损失并不一定是最优选择. 在更复杂的 NLSE、

长距离传播问题中, PDE 残差、初边值约束与观测数据项可能存在明显量级差异, 从而导致梯度竞争或优化停滞. 后续可引入基于梯度范数、残差大小或不确定性估计的自适应权重策略, 以进一步提升训练稳定性. 其二, 孤子演化问题的计算区域中通常存在较大范围的低幅值或缓变区域, 后续可引入自适应采样策略, 将更多配置点分配至孤子主体及误差集中区域, 从而进一步提升训练效率与求解精度.

感谢北京邮电大学物理科学与技术学院刘慕威博士的讨论.

附录 A

表 A1 参数退化情形下的实验参数设置

Table A1. Experimental settings for the degenerate-parameter case.

参数项	PI-CVNN	PINN	PINN-param
网络层数	5	5	5
网络宽度	40	40	40
参数量	10162	5122	11282
优化算法	Adam+L-BFGS	Adam+L-BFGS	Adam+L-BFGS
Adam训练步数	1000	1000	1000
L-BFGS最大迭代次数	2000	2000	2000
计算区域 (t, x)	$[-8, 3] \times [0, 1]$	$[-8, 3] \times [0, 1]$	$[-8, 3] \times [0, 1]$
配置点数 N_f	5000	5000	5000
初值点数 N_{ic}	50	50	50
边界点数 N_{bc}	50×2	50×2	50×2
随机种子	{0, 1, 2, 3, 4}	{0, 1, 2, 3, 4}	{0, 1, 2, 3, 4}
数据精度	float32	float32	float32
GPU	RTX 3060 12 GB	RTX 3060 12 GB	RTX 3060 12 GB

表 A2 参数退化情形下的详细求解结果

Table A2. Detailed results for the degenerate-parameter case.

模型	指标	均值	标准差
PINN	相对 L_2 误差	1.421×10^{-3}	2.200×10^{-4}
PINN	振幅误差	1.023×10^{-3}	1.475×10^{-4}
PINN	相位误差	6.738×10^{-3}	1.554×10^{-3}
PINN	最终损失	1.923×10^{-7}	1.406×10^{-7}
PINN	训练时间/s	213	
PINN-param	相对 L_2 误差	1.254×10^{-3}	2.389×10^{-4}
PINN-param	振幅误差	9.273×10^{-4}	2.144×10^{-4}
PINN-param	相位误差	6.356×10^{-3}	1.216×10^{-3}
PINN-param	最终损失	1.152×10^{-7}	2.630×10^{-8}
PINN-param	训练时间/s	329	
PI-CVNN	相对 L_2 误差	1.278×10^{-3}	1.631×10^{-4}
PI-CVNN	振幅误差	9.160×10^{-4}	1.274×10^{-4}
PI-CVNN	相位误差	6.869×10^{-3}	1.417×10^{-3}
PI-CVNN	最终损失	1.035×10^{-7}	4.902×10^{-8}
PI-CVNN	训练时间/s	814	

表 A3 常系数高阶情形下的实验参数设置

Table A3. Experimental settings for the constant-coefficient higher-order case.

参数项	PI-CVNN	PINN	PINN-param
网络层数	5	5	5
网络宽度	80	80	114
参数量	39522	19842	39902
优化算法	Adam+L-BFGS	Adam+L-BFGS	Adam+L-BFGS
Adam训练步数	1000	1000	1000
L-BFGS最大迭代次数	2000	2000	2000
计算区域 (t, x)	$[-10, 5] \times [0, 2]$	$[-10, 5] \times [0, 2]$	$[-10, 5] \times [0, 2]$
配置点数 N_f	5000	5000	5000
初值点数 N_{ic}	50	50	50
边界点数 N_{bc}	50×2	50×2	50×2
随机种子	{0, 1}	{0, 1}	{0, 1}
数据精度	float32	float32	float32
GPU	RTX 3060 12 GB	RTX 3060 12 GB	RTX 3060 12 GB

表 A4 常系数高阶情形下的详细求解结果

Table A4. Detailed results for the constant-coefficient higher-order case.

模型	指标	均值	标准差
PINN	相对 L_2 误差	2.262×10^{-3}	4.744×10^{-4}
PINN	振幅误差	1.562×10^{-3}	1.877×10^{-4}
PINN	相位误差	1.847×10^{-2}	5.586×10^{-3}
PINN	最终损失	8.488×10^{-8}	4.399×10^{-8}
PINN	参数量	19842	0
PINN	训练时间/s	251	
PINN-param	相对 L_2 误差	3.080×10^{-3}	2.628×10^{-4}
PINN-param	振幅误差	2.237×10^{-3}	2.621×10^{-4}
PINN-param	相位误差	3.395×10^{-2}	1.227×10^{-2}
PINN-param	最终损失	2.735×10^{-7}	6.939×10^{-8}
PINN-param	参数量	39902	0
PINN-param	训练时间/s	327	
PI-CVNN	相对 L_2 误差	2.136×10^{-3}	1.075×10^{-3}
PI-CVNN	振幅误差	1.674×10^{-3}	7.289×10^{-4}
PI-CVNN	相位误差	2.326×10^{-2}	1.749×10^{-2}
PI-CVNN	最终损失	7.597×10^{-8}	2.358×10^{-8}
PI-CVNN	参数量	39522	0
PI-CVNN	训练时间/s	1134	

参考文献

- [1] Zabusky N J, Kruskal M D 1965 *Phys. Rev. Lett.* **15** 240
- [2] Costa A, Osborne A R, Resio D T, Alessio S, Chirivì E, Saggese E, Bellomo K, Long C E 2014 *Phys. Rev. Lett.* **113** 108501
- [3] Zhu Y, Semisalov B, Krstulovic G, Nazarenko S 2023 *Phys. Rev. Lett.* **130** 133001
- [4] Heidemann R, Zhdanov S K, Sütterlin R, Thomas H M, Morfill G E 2009 *Phys. Rev. Lett.* **102** 135002
- [5] Wen J M, Bo W B, Wen X K, Dai C Q 2023 *Acta. Phys. Sin.* **72** 100502 (in Chinese) [温嘉美, 薄文博, 温学坤, 戴朝卿 2023 物理学报 **72** 100502]
- [6] Wang J F, Wei X, Liu S, Yang L Z, Xue P P, Fan L L 2024

- Acta. Phys. Sin.* **73** 094205 (in Chinese) [王娟芬, 韦鑫, 刘帅, 杨玲珍, 薛萍萍, 樊林林 2024 物理学报 **73** 094205]
- [7] Hasegawa A, Tappert F 1973 *Appl. Phys. Lett.* **23** 142
- [8] Wang T Y, Zhou Q, Liu W J 2022 *Chin. Phys. B* **31** 020501
- [9] Wang D S, Jiang X T, Song Y C, Fu M X, Zhang Z G, Chen X, Zhang M 2022 *IEEE Commun. Mag.* **60** 32
- [10] Zhang Z Y, Liu M W, Liu W J 2025 *J. Syst. Sci. Complex.* **38** 2732
- [11] Tian S F, Yu Y X, Li B 2025 *Chin. Phys. Lett.* **42** 120001
- [12] Zhou H S, Zhang Z Y, Liu M W, Qi F H, Liu W J 2025 *Chin. Phys. Lett.* **42** 070001
- [13] Tian S F, Li B, Zhang Z 2024 *Chin. Phys. Lett.* **41** 030201
- [14] Hu C, Zhang Z Z, Liu M W, Xiang L Y, Wu H J, Liu W J, He Z F 2024 *Chin. Phys. Lett.* **41** 120303
- [15] Raissi M, Perdikaris P, Karniadakis G E 2019 *J. Comput. Phys.* **378** 686
- [16] Cuomo S, Schiano Di Cola V, Giampaolo F, Rozza G, Raissi M, Piccialli F 2022 *J. Sci. Comput.* **92** 88
- [17] Karniadakis G E, Kevrekidis I G, Lu L, Perdikaris P, Wang S F, Yang L 2021 *Nat. Rev. Phys.* **3** 422
- [18] Li J, Chen Y 2020 *Commun. Theor. Phys.* **72** 105005
- [19] Li J, Chen Y 2020 *Commun. Theor. Phys.* **72** 115003
- [20] Pu J C, Li J, Chen Y 2021 *Chin. Phys. B* **30** 060202
- [21] Pu J C, Li J, Chen Y 2021 *Nonlinear Dyn.* **105** 1723
- [22] Pu J C, Peng W Q, Chen Y 2021 *Wave Motion* **107** 102823
- [23] Li J, Chen Y 2021 *Commun. Theor. Phys.* **73** 015001
- [24] Lin S N, Chen Y 2022 *J. Comput. Phys.* **457** 111053
- [25] Lee C Y, Hasegawa H, Gao S 2022 *IEEE/CAA J. Autom. Sin.* **9** 1406
- [26] Zhang L, Du M G, Bai X D, Chen Y T, Zhang D X 2025 *Phys. Rev. Res.* **7** 013164
- [27] Agrawal G P 2013 *Nonlinear Fiber Optics* (5th Ed.) (Amsterdam: Academic Press) pp10–70
- [28] Hirota R 1973 *J. Math. Phys.* **14** 805

Complex neural network solution of the nonlinear Schrödinger equation in fiber optic communication systems*

WANG Libin ZHANG Zhiyang LIU Wenjun[†]

(School of Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

(Received 10 April 2026; revised manuscript received 20 May 2026)

Abstract

The nonlinear Schrödinger equation is a fundamental model for describing the propagation of wave packets in dispersive and nonlinear media. It is widely used in nonlinear optics, plasma physics, and other fields that involve complex wave dynamics. This paper proposes PI-CVNN, a physics-informed complex-valued neural network for solving the nonlinear Schrödinger equation. The proposed method directly models the optical envelope in the complex domain and incorporates the initial condition, boundary condition, and governing-equation residual into a unified physics-informed loss function. By using complex-valued weight matrices, PI-CVNN explicitly represents the coupling between the real and imaginary parts, thereby better characterizing the evolution of the complex envelope. Experimental results show that PI-CVNN effectively approximates the complex-valued dynamics of optical soliton propagation in fibers. This study provides a useful tool for modeling nonlinear effects in optical fiber communications and broadens the application potential of complex-valued neural networks in scientific computing.

Keywords: nonlinear Schrödinger equation, optical solitons, physics-informed neural networks, complex-valued neural networks

DOI: 10.7498/aps.75.20260506

CSTR: 32037.14.aps.75.20260506

* Project supported by the Key Program of the National Natural Science Foundation of China (Grant No. 2022YFB4601101), the National Natural Science Foundation of China (Grant No. 12261131495), and the Beijing University of Posts and Telecommunications Foundation, China (Grant No. 2025JCTP01).

[†] Corresponding author. E-mail: junglin@bupt.edu.cn



光纤通信系统中非线性薛定谔方程的复数神经网络求解

王立斌 张之阳 刘文军

Complex neural network solution of the nonlinear Schrödinger equation in fiber optic communication systems

WANG Libin ZHANG Zhiyang LIU Wenjun

引用信息 Citation: *Acta Physica Sinica*, 75, 190001 (2026) DOI: 10.7498/aps.75.20260506

CSTR: 32037.14.aps.75.20260506

在线阅读 View online: <https://doi.org/10.7498/aps.75.20260506>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于扩展混合训练物理信息神经网络的非线性薛定谔方程求解和参数发现

Solving nonlinear Schrödinger equations and parameter discovery via extended mixed-training physics-informed neural networks

物理学报. 2025, 74(16): 160201 <https://doi.org/10.7498/aps.74.20250422>

基于梯度优化物理信息神经网络求解复杂非线性问题

Solving complex nonlinear problems based on gradient-optimized physics-informed neural networks

物理学报. 2023, 72(10): 100202 <https://doi.org/10.7498/aps.72.20222381>

信息表征-损失优化改进的物理信息神经网络求解偏微分方程

allaPINNs: A physics-informed neural network with improvement of information representation and loss optimization for solving partial differential equations

物理学报. 2025, 74(18): 188701 <https://doi.org/10.7498/aps.74.20250707>

基于物理信息神经网络的绝热捷径动力学分析

Dynamic analysis of shortcut to adiabaticity based on physical information neural network

物理学报. 2025, 74(11): 114204 <https://doi.org/10.7498/aps.74.20250147>

基于物理信息神经网络求解低温等离子体反问题

Solving inverse problems of low-temperature plasmas by physics-informed neural networks

物理学报. 2025, 74(23): 235207 <https://doi.org/10.7498/aps.74.20251159>

基于源项解耦的物理信息神经网络方法及其在放电等离子体模拟中的应用

Physics-informed neural networks based on source term decoupled and its application in discharge plasma simulation

物理学报. 2024, 73(14): 145201 <https://doi.org/10.7498/aps.73.20240343>