



Pseudo grid-based physics-informed convolutional-recurrent network solving the integrable nonlinear lattice equations

Zhe Lin^a, Yong Chen^{a,b,*}

^a School of Mathematical Sciences, Key Laboratory of Mathematics and Engineering Applications (Ministry of Education) and Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China

^b College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China

ARTICLE INFO

Communicated by Boumediene Hamzi

Keywords:

Pseudo grid-based physics-informed convolutional-recurrent network
Integrable nonlinear lattice equations
Multi-soliton solutions
Rational soliton solutions

ABSTRACT

Traditional discrete learning methods involve discretizing continuous equations using difference schemes, necessitating considerations of stability and convergence. Integrable nonlinear lattice equations possess a profound mathematical structure that enables them to revert to continuous integrable equations in the continuous limit, particularly retaining integrable properties such as conservation laws, Hamiltonian structure, and multiple soliton solutions. The pseudo grid-based physics-informed convolutional-recurrent network (PG-PhyCRNet) is proposed to investigate the localized wave solutions of integrable lattice equations, which significantly enhances the model's extrapolation capability to lattice points beyond the temporal domain. We conduct a comparative analysis of PG-PhyCRNet with and without pseudo grid by investigating the multi-soliton solutions and rational solitons of the Toda lattice and self-dual network equation. The results indicate that the PG-PhyCRNet excels in capturing long-term evolution and enhances the model's extrapolation capability for solitons, particularly those with steep waveforms and high wave speeds. Finally, the robustness of the PG-PhyCRNet method and its effect on the prediction of solutions in different scenarios are confirmed through repeated experiments involving pseudo grid partitioning.

1. Introduction

Solving systems of continuous and discrete nonlinear partial differential equations (PDEs) has been a key component in scientific computing and mathematical physics. In recent years, deep learning has achieved great success in solving PDEs due to the powerful representation capabilities of neural networks [1–3]. Research on modeling mathematical-physical systems using deep neural networks can be broadly classified into two categories: continuous networks and discrete networks. Both continuous and discrete integrable systems, a special class of nonlinear systems, possess a rich mathematical structure and unique properties, such as Lax pairs, Hamiltonian structures, and infinite conservation laws, distinguishing them from ordinary nonlinear systems [4–6].

1.1. Continuous and discrete learning methods

Significant progress has been made in the study of nonlinear system surrogate modeling using deep learning [2,3,7,8]. Many studies focus on minimizing the physical information error function, driving the

vibrant growth of scientific machine learning research, particularly in the modeling and simulation of PDEs [1,9]. Karniadakis' group proposed a landmark data-driven approach based on fully connected neural networks (FCNs): Physics-Informed Neural Networks (PINNs), which is an efficient and powerful method for solving the forward and inverse problems of PDEs [1]. Compared to purely data-driven neural network learning, PINN imposes physical information constraints during training, allowing for the learning of more generalizable models with fewer data samples and the wide application in aerodynamic flows, solid mechanics, blood flows modeling and many others [10–13]. Chen's group proposed the concept and framework of integrable deep learning and conducted a series of significant studies. Many data-driven localized wave solutions of continuous integrable equations can be obtained by PINN [14–19]. Besides, they investigated some complex problems in integrable systems using the improved PINN method: high-dimensional integrable systems, nonlocal integrable equations, the forward and inverse problem for variable coefficients integrable equations, and solving the higher-order RW solutions [20–23]. There are also some innovative works that integrate the perfect

* Corresponding author at: School of Mathematical Sciences, Key Laboratory of Mathematics and Engineering Applications (Ministry of Education) and Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China.

E-mail address: ychen@sei.ecnu.edu.cn (Y. Chen).

<https://doi.org/10.1016/j.physd.2024.134304>

Received 26 June 2024; Accepted 20 July 2024

Available online 30 July 2024

0167-2789/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

mathematical structures of integrable systems. For example, the two-stage PINN based on conservation laws can enhance the prediction accuracy and generalization ability of solutions [24], the PINN based on the Miura transformation discovered a new kink-bell solution of modified Korteweg de-Vries (mKdV) equation [25], and the LPNN based on Lax pairs accelerated training efficiency and improves the prediction accuracy of localized wave solutions [26]. Additionally, some other excellent papers on integrable deep learning that readers can refer to Refs. [27–30].

Unlike FCNs, convolutional neural networks (CNNs) extract spatial hierarchical features through local connections and weight sharing, significantly reducing the number of parameters and computational complexity. This makes CNNs more efficient for handling large-scale data and high-dimensional problems, and particularly well-suited for surrogate modeling of discretized PDEs. Discretized learning approaches based on CNNs have demonstrated good convergence and scalability in modeling PDEs [31–38]. Researchers applied CNNs to surrogate modeling and uncertainty quantification of PDEs in rectangular reference domains [31]. PhyGeoNet achieved geometrically adaptive solutions for steady-state partial differential equations through coordinate transformations [32]. The AR-DenseED method successfully solved PDEs using discretized learning without labeled data [33]. The physics-informed convolutional-recurrent network (PhyCRNet) method, based on an encoder-decoder convolutional long short-term memory (LSTM) network, inherited the advantages of AR-DenseED in feature extraction in low-dimensional space and time evolution learning, demonstrating superior performance in the accuracy, extrapolation, and generalization of PDEs' solutions [38].

1.2. Discrete integrable systems and deep learning

The pioneering study on nonlinear lattice equations (NLEs) can be traced back to the early 1950s, when Fermi, Pasta, and Ulam (FPU) conducted great numerical experiments [39]. They established a model of vibrating particles connected by nonlinear springs and, through computer simulations, discovered that waves with various sinusoidal initial conditions would return to their initial energy states after a certain period. To extract quantitative information from models described by continuous differential equations, it is often necessary to solve them numerically using various discretization methods. However, it is generally expected that the discretization of any given integrable equations may lead to a non-integrable system. It is crucial to assure that the discretized models exhibit the same qualitative features of the dynamics as their continuous counterparts. Therefore, Toda investigated the motion equations of a one-dimensional uniform lattice with nearest-neighbor interactions and propose the integrable Toda lattice [5,6,40–42]. The existence of analytical solutions in integrable lattice equations opens up new avenues for a better understanding of nonlinear phenomena, not only in the context of FPU recurrence but also in other areas related to nonlinear mechanics [43–45]. Until the 1970s, the topic of integrable discretization continued to advance. Case and Kac [46] focused on the discretization of the Schrödinger spectral problem and the discretization of the Ablowitz-Kaup-Newell-Segur spectral problem (developed by Ablowitz and Ladik [6,43]) as well as establishing a difference-based inverse scattering transformation. Hirota [47,48] utilized the bilinear Bäcklund transformations and their connection with Lax pairs to derive a series of discrete integrable systems for bilinear equations. Subsequent research on NLEs have expanded, including the nonlinear self-dual network equation under nonlinear inductor-capacitor (LC) circuits [49,50], Ablowitz-Ladik (AL) lattice [6,43], Błaszak-Marciniak lattice [51], Volterra lattice [52], etc. These equations have made significant theoretical contributions to describing physical phenomena such as crystalline phenomena and nonlinear lattice dynamics [53,54]. Methods for solving these lattice equations analytically, such as the inverse scattering transformation method [43], Bäcklund transformation [55], Hirota method [56,57],

and Darboux transformation [58], have been developed to better understand these mathematical structures of lattice solutions and physical phenomena.

Discrete integrable systems share many similarities with their continuous counterparts, which possess a pair of linear spectral problems (known as discrete Lax pairs), infinite conservation law, Hamiltonian structure, symmetries, and other features. In this paper, some significant discrete integrable systems can converge to their corresponding continuous integrable systems by taking the appropriate continuum limit:

- The first equation discussed in this paper is the Toda lattice [5]:

$$\frac{d^2}{dt^2} (\ln u_n) = u_{n+1} + u_{n-1} - 2u_n, \quad (1)$$

or another coupled form:

$$\dot{\mathbf{p}}_n = \begin{pmatrix} v_n - v_{n-1} \\ u_{n+1} - u_n \end{pmatrix}, \quad (2)$$

where $\mathbf{p}_n = (\ln u_n, v_n)^T$, $\dot{\mathbf{p}}_n = d\mathbf{p}_n/dt$. Inspired by Refs. [47,48,59], we give a step-size parameter δ and transformation

$$\begin{aligned} u_n(t) &= 1 + \delta q (\delta n - 6\delta t, \delta^2 t) := 1 + \delta q(x, \tau), \\ v_n(t) &= 1 + \delta p \left(\delta n + \frac{\delta^2}{4} t, \delta t \right) := 1 + \delta p(y, k). \end{aligned} \quad (3)$$

Eliminating the linear terms in the system of equations, the lattice (2) can approximate:

$$(q_\tau + q_{xxx} + 6qq_x)\delta^3 + \mathcal{O}(\delta^4) = 0, \quad (4)$$

which is just the famous Korteweg de-Vries (KdV) equation when the higher-order terms are neglected. Eq. (4) was first derived by Korteweg and de-Vries in the study of long water waves in relatively shallow channels, which has important applications in nonlinear science and mathematical physics [60,61].

- The second equation discussed is self-dual network equation [62]:

$$\dot{\mathbf{y}}_n = \begin{bmatrix} (1 + I_n^2)(V_{n-1} - V_n) \\ (1 + V_n^2)(I_n - I_{n+1}) \end{bmatrix}, \quad (5)$$

where $\mathbf{y}_n = (I_n, V_n)^T$. We give the transformation

$$I_n(t) = \alpha + \Delta(\alpha)q(x, \tau), \quad V_n(t) = \alpha + \Delta(\alpha)p(x, \tau), \quad (6)$$

with

$$\Delta(\alpha) = \frac{(\alpha^2 + 1)\delta}{\sqrt{1 - 3\alpha^2}}, \quad x = \delta n + b\delta t, \quad \tau = -\frac{1}{6}b\delta^3 t, \quad 0 < \alpha < 1, b \in \mathbb{R},$$

the lattice (5) can approximate:

$$(q_\tau + q_{xxx} + 6q^2q_x)\delta^4 + \mathcal{O}(\delta^5) = 0, \quad (7)$$

which is just the mKdV when the higher-order terms are neglected. The mKdV equation was first introduced by Miura and can serve as a model equation for acoustic waves in anharmonic lattices [63]. It can also describe dust-acoustic solitary waves in dusty plasmas and wave phenomena in nonlinear optics [64,65].

The step-size parameter δ measures the grid size of the spatial component and allows us to bring lattice points closer together to create a continuous point. It is not difficult to observe that higher-order dispersion terms in continuous integrable equations transform into a simple translation relationship under integrable discretization. This greatly facilitates our numerical computations. From this perspective, we are considering whether discrete learning methods in deep learning can be used to further investigate the dynamical behavior of solutions in the integrable lattice equations.

To the best of our knowledge, researches on using deep learning to solve problems related to NLEs are relatively limited. Saqlain et al. have addressed the inverse problem of parameter identification in

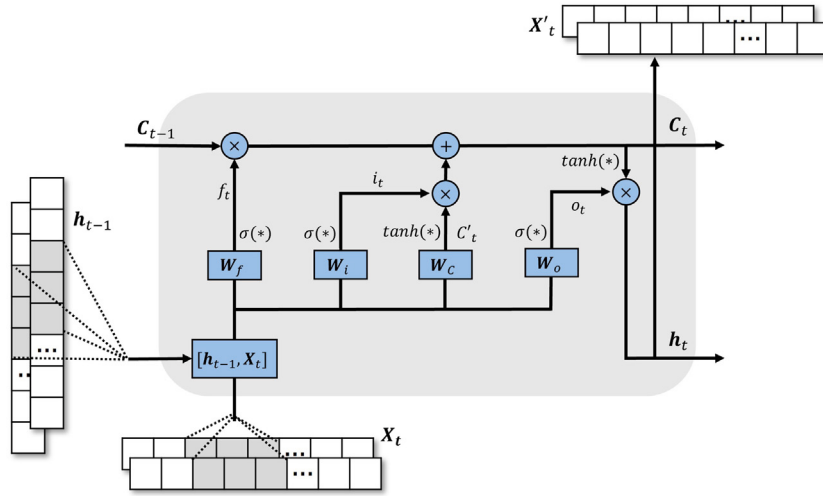


Fig. 1. Chain-like structure of ConvLSTM.

discrete high-dimensional systems by appropriately adjusting the PINN method [66,67]. In data-driven mainstream learning processes, interpolation and extrapolation are involved in evaluating model performance. Interpolation requires the generalization ability within the boundaries of the training data. Extrapolation refers to estimating unknown values by extending beyond the known range based on known sequences or factors [68]. PINN learns the function mappings directly in spatio-temporal regions, which leads to a lack of spatiotemporal extrapolation ability due to not focusing on the temporal evolution of physical processes. However, the discrete learning method of PhyCRNet discretizes PDEs and establishes dependencies between temporal data through CNNs and LSTM, ensuring that the model solution possesses a certain degree of extrapolation ability in the temporal direction. Besides, CNNs focus on the discrete points of the input tensor, making them more suitable for exploring discrete systems. Compared to continuous learning methods like PINN, we prefer to use discrete learning methods that directly target the discrete grid points in lattice equations to explore the dynamical behavior of the solutions. Enhancing extrapolation is often achieved by altering the model structure or providing more prior conditions. The pseudo grid-based physics-informed convolutional-recurrent network (PG-PhyCRNet) method, which incorporates pseudo grid training, is based on the richer prior conditions (initial value conditions) in the theory of discrete integrable systems. It addresses the issue of irregular distribution of integer lattice, which causes the network to fail to capture the potential function features and leads to prediction failures. Thus, the motivation and highlights of this paper are:

- Our motivation is to utilize PhyCRNet methods to directly explore integrable NLEs, which is a novel attempt. In the traditional PhyCRNet learning methods, it is essential to carefully consider the stability and convergence of the discretization of continuity equations using finite difference schemes (especially for higher-order dispersion and nonlinear terms). However, as a unique discrete scheme of its integrable continuous equations, NLEs possess a richer mathematical structure and retains the same integrable properties. Moreover, under the continuous limit (3) and (6), this may allow an indirect discussion of the dynamical behavior of continuous solutions with the help of discrete solutions.
- PG-PhyCRNet significantly enhances the model's extrapolation capability to lattice points beyond the temporal domain. Based on more prior information about initial values obtained by the integrable nature of discrete lattice equations, PG-PhyCRNet enables easier to learn the dynamical features of lattice solitons.
- The PG-PhyCRNet approach offers superior performance in capturing the dynamics of high-speed, steep-profile solitons, particularly in extrapolating solutions beyond the initial time domain.

Multi-soliton and rational soliton solutions of the Toda lattice and self-dual network equations are investigated by using the PhyCRNet method with and without pseudo grid (PG-PhyCRNet and PhyCRNet). The performance of these two training methods in different scenarios are comparatively analyzed: the PhyCRNet method is better suited for lattice solitons with lower speeds and smoother profiles; the PG-PhyCRNet method is better suited for lattice solitons with higher speeds and steeper profiles, where training with a pseudo grid gives an accuracy improvement in the prediction of the lattice solution outside the time domain.

The structure of this paper is as follows: Section 2 provides a model description and details PG-PhyCRNet method. Sections 3 and 4 discuss the numerical experiments of the PhyCRNet method without and (PG-PhyCRNet and PhyCRNet) in predicting different types of lattice solitons, respectively. In Section 5, the effect of pseudo grid in PG-PhyCRNet method and the stability of these localized waves are analyzed through repeated experiments. Finally, the last section provides a discussion and summary of this paper.

2. Model description

In this section, we will introduce the PG-PhyCRNet for solving $(1+1)$ -dimensional semi-discrete integrable systems. The architecture's ability to handle different solutions of semi-discrete integrable systems from two perspectives are discussed. Both perspectives involve extracting low-dimensional spatial features and learning the temporal evolution of compressed information through an encoder-decoder convolutional-recurrent scheme.

2.1. ConvLSTM

ConvLSTM is an architecture that combines CNNs and LSTMs specifically designed for processing sequential data [69,70]. Unlike traditional LSTMs, ConvLSTM applies convolution operations at each time step, which helps capture spatial information in the sequential data [70]. It consists mainly of input gates, forget gates, output gates, and convolutional layers. In the gating mechanism, ConvLSTM uses convolutional layers to compute the relevance between the input at the current time step and the hidden state from the previous time step, which is used to control the input and output information at the current time step and to forget the information from the previous hidden state. Since the paper exclusively addresses $(1+1)$ -dimensional NLEs, the convolution operations in the convolutional layers of the ConvLSTM are all one-dimensional. For convenience, we uniformly refer to the 1D-ConvLSTM as ConvLSTM.

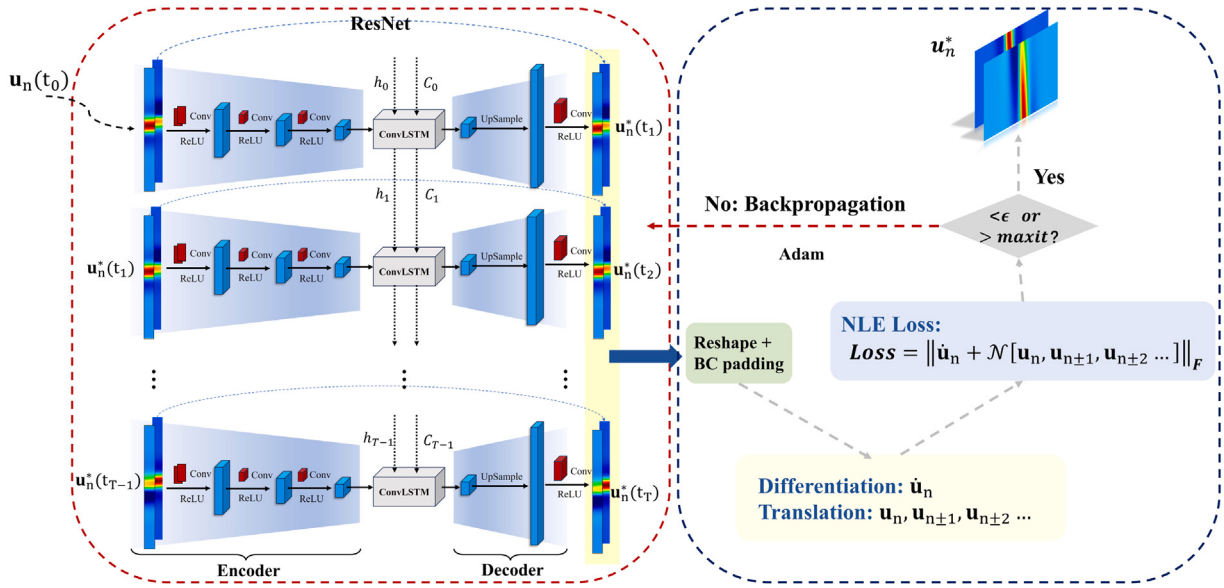


Fig. 2. Network architecture. “Reshape” represents the reshaping of the potential function tensor at each epoch. “BC padding” represents the boundary condition.

ConvLSTM also implements three gates (the forget gate f_t , input gate i_t , and output gate o_t) to protect and control the cell state. Let X_t be the input tensor, and $\{h_{t-1}, C_t\}$ be the hidden state and cell state to be updated at time step t , as shown in the chain-like structure in Fig. 1. The forgetting gate f_t reads the information from $\{h_{t-1}, X_t\}$ and passes through the sigmoid layer $\sigma(\cdot)$ to obtain a value between 0 and 1. When the forget gate is close to 1, the information in the cell state is fully retained; otherwise, the information is completely forgotten. This adaptive behavior allows the forget gate to clear memorized information in the cell state C_{t-1} . The input gate i_t has two parts: the first sigmoid layer decides what to update, and the second tanh layer updates the information into the cell state. The output gate o_t uses a sigmoid layer determines how much of the current cell state to output. The mathematical formula for the ConvLSTM unit can be expressed as:

$$\begin{aligned} i_t &= \sigma(W_i * [X_t, h_{t-1}] + b_i), \quad f_t = \sigma(W_f * [X_t, h_{t-1}] + b_f), \\ \tilde{C}_{t-1} &= \tanh(W_c * [X_t, h_{t-1}] + b_c), \quad C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_{t-1}, \\ o_t &= \sigma(W_o * [X_t, h_{t-1}] + b_o), \quad h_t = o_t \odot \tanh(C_t), \end{aligned} \quad (8)$$

where $*$ is the convolutional operation and $\odot$ denotes the Hadamard product. The presence of ConvLSTM ensures that in the process of solving integrable lattice equations, the dependence relationship between the potential function and the preceding and succeeding states is preserved. Moreover, the information flowing into ConvLSTM is compressed and latent features are extracted by an encoder.

2.2. Network architecture

PhyCRNet is a discrete learning method for solving multi-dimensional spatiotemporal PDEs without any labeled data [38]. The architecture mainly includes an encoder–decoder structure, residual connections, autoregressive processes, and differentiation based on gradient-free convolutional kernels. The exploration of integer lattice points in integrable NLEs is essential. Therefore, the initial consideration is to directly apply PhyCRNet to train and predict integer lattice points. However, due to the irregular distribution of many lattice solitons near the wave peaks, more prior information needs to be incorporated into the model for joint training. This approach can lead to better predictions for integer lattice points. Therefore, we propose the PG-PhyCRNet method to solve the integrable lattice equation in two different situation:

- We prefer to focus on the dynamical behavior of integer lattice sites in the spatial part of the NLE. Hence, we illustrate this approach using the initial-boundary value problem of (1+1)-dimensional NLE as an example:

$$\begin{cases} \dot{u}_n + \mathcal{N}[u_n, u_{n\pm 1}, u_{n\pm 2} \dots] = 0, \\ u_n(0) = u(n, 0), \quad n \in \Omega, \\ u_{n_1} = u_{b_1}, \quad u_{n_N} = u_{b_2}, \quad t \in [t_0, t_T], \end{cases} \quad (9)$$

where $u_n = u(n, t)$ denotes the solution of NLE, $\mathcal{N}$ is the nonlinear operator, $n \in \mathbb{Z}$ stands for lattice site and the region $\Omega = [n_1, n_N]$ is a finite one-dimension lattice, u_{b_1} and u_{b_2} are constants.

- If the additional information beyond the initial values of the integer lattice sites can be obtained, such as the initial value samples of the integer lattice sites following a certain distribution or empirical regression function (i.e., the initial condition in Eq. (9) are changed to $u_n(0) = u(n, 0)$, $n \in \Omega'$, $\Omega' = [n_1, n_1 + i \frac{n_N - n_1}{N - 1}]$, $i = 1, \dots, N - 1$), then the trend of integer lattice sites with help with non-integer grid can be studied. We refer to these non-integer grid points as “pseudo grid” of the network. This operation is actually to better fit the continuous potential function $u(n, t)$ with respect to variables n and t through model training. For cases like solitary waves with fast wave speeds and narrow wave widths, it is often difficult to capture their propagation characteristics near the peak using integer lattice sites. The addition of pseudo grid avoids this problem and enables the solution u_n based on the well-trained model to have good extrapolation ability in the time direction. However, the drawback of this method is that it may lose prediction accuracy within the given region.

By utilizing the exact solutions of discrete integrable systems to provide rich training samples, both training methods under different prior conditions can be satisfied. Fig. 2 illustrates the PG-PhyCRNet architecture for solving discrete integrable systems. The encoder structure consists of three convolutional operations, which are used to learn low-dimensional latent features from the input state variables $u_n(t_i)$ ($i = 0, 1, \dots$), with a ReLU function applied after each convolution for non-linearity. The ConvLSTM layer receives the feature information passed by the encoder and propagates the temporal feature information through the latent features h_t and cell states C_t . The decoder includes a pixel shuffle operation [71] and a convolution operation with a large kernel size, aiming to scale the low-resolution feature information

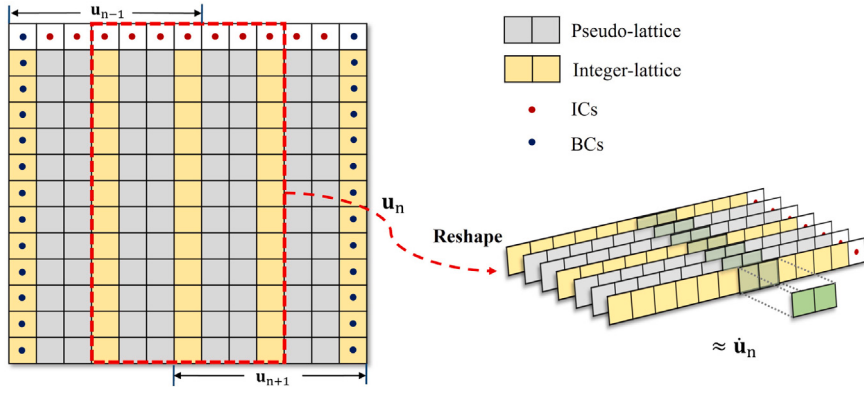


Fig. 3. The translation operations in the spatial direction (left) and the discrete numerical differentiation in the time direction (right).

passed by the ConvLSTM layer back to the original size of interest and reconstruct it as the information for the next time step in high resolution. Finally, a simple autoregressive process is achieved by using residual connections to update the potential function states for adjacent time steps, where $u_n(t_{i+1}) = u_n(t_i) + \delta_t * \mathcal{N}\mathcal{N}[\cdot]$, where $\mathcal{N}\mathcal{N}$ denotes the trained network operator and δ_t is the time interval.

2.3. Translation operation and derivative approximation

After obtaining the output from the model on the left side of Fig. 2, the outputs are concatenated at each time step and the boundary conditions (BCs) is hard imposed. This is done to facilitate the subsequent use of translation operation and gradient-free convolution filters to represent the translation operations ($E^k f_n = f_{n+k}$, $n, k \in \mathbb{Z}$) in the spatial direction and discrete numerical differentiation in the time direction.

Due to the higher-order dispersion terms in continuous integrable equations can transform into a simple translation relationship under integrable discretization, the translation operator is very important as a surrogate in the network. We give the following two ways of handling the translation operator based on model training methods without and with pseudo grid:

(1) Translated potential function $u_{n\pm 1}, u_{n\pm 2}, \dots$ can be obtained by directly shifting the concatenated predicted values. It is worth noting that for the first training method without adding pseudo grid, we simply shift the predicted solution forward or backward by k steps to obtain $u_{n\pm k}$.

(2) For the training method with pseudo grid, we need to shift the predicted solution forward or backward by k/δ_n ($\delta_n \in \mathcal{A}$) steps to obtain $u_{n\pm k}$, where

$$\mathcal{A} = \left\{ \delta_n \mid \delta_n = \frac{n_N - n_1}{N_n - 1}, N_n \geq 2, N_n \in \mathbb{Z}_+, \frac{1}{\delta_n} \in \mathbb{Z}_+ \right\}. \quad (10)$$

The corresponding processes of tensor recombination, hard-imposed of BCs, and translation operators are illustrated in Fig. 3. The left side in Fig. 3 shows the translation operation in the training method with pseudo grid, where the red dashed part represents the potential function u_n that will enter the loss function. And the translation operator in NLEs can be replaced by a proper slicing operation on the output.

Besides, we use forward differential format from the finite difference method to approximate the derivatives in the time direction. From the right side of Fig. 3, it can be seen that we decompose the two-dimensional potential function into a one-dimensional potential function that varies with time. We approximate the derivative in the time direction by applying a one-dimensional gradient-free convolution kernel of size 2 to the reorganized potential function.

2.4. Parameter optimization and capability evaluation of PG-PhyCRNet

The essence of training a neural network or deep learning model lies in updating its weights and biases. Based on the training data, our goal is to minimize the value of the loss function by optimizing the parameters of the neural network. The presence of the loss function indicates the direction of training for deep learning models.

Due to the imposition of strict initial boundary conditions, there is no longer a need to include these conditions as constraints in the loss function. Therefore, the loss function only needs to define the required physical information to be satisfied. Building upon the two training methods mentioned in the previous section, we uniformly define their residual $\mathcal{R}(\mathbf{n}, t; \theta)$ as:

$$\mathcal{R}(\mathbf{n}, t; \theta) = \dot{u}_n + \mathcal{N}[u_n, u_{n\pm 1}, u_{n\pm 2}, \dots]. \quad (11)$$

This expression actually represents the left-hand side of Eq. (9), where $\theta = \{\mathbf{w}, \mathbf{b}\}$ denotes the parameters to be optimized in the neural network. Building upon the two training methods mentioned in the previous subsection, we define the following two loss functions:

$$\begin{aligned} \text{MSE}_1(\theta) &= \frac{1}{N_T N} \sum_{i=0}^{N_T-1} \sum_{j=0}^{N-1} \left| \mathcal{R}(n_{f_1}^j, t_{f_1}^j; \theta) \right|^2, \\ \text{MSE}_2(\theta) &= \frac{1}{N_T N_n} \sum_{i=0}^{N_T-1} \sum_{j=0}^{N_n-1} \left| \mathcal{R}(n_{f_2}^j, t_{f_2}^j; \theta) \right|^2, \end{aligned} \quad (12)$$

where the independent variables $n_{f_1}^j \in \Omega$, $n_{f_2}^j \in \Omega'$, and $t_{f_1}^j \in [t_0, t_0 + i \frac{t_T - t_0}{N_T - 1}]$ ($i = 1, 2, \dots, N_T - 1$). By minimizing the mean squared error criterion, we optimize the parameters of the neural network to satisfy the provided physical information. We choose to use the Adam optimization algorithm to optimize the loss function. Ultimately, the trained PG-PhyCRNet model can not only provide numerical solutions within the specified region but also forecast and provide numerical solutions outside the time domain except for the boundary values.

We need to use the exact solution of the physical equations to evaluate the application performance of PG-PhyCRNet in discrete integrable systems and assess the propagation error through two stages: training and extrapolation. Although the use of PhyCRNet involves the above two methods to meet different needs, the ultimate goal is to evaluate the predictive accuracy of the model on integer grid sites. The first stage involves evaluating the model's training generalization ability through the relative $\mathbb{L}_2$ error of $N \times N_t$ ($N_t = (N_T + 1)/2$) data points on the given computational grid $\Omega \times [t_0, t_0 + i \frac{t_T - t_0}{N_t - 1}]$ ($i = 1, 2, \dots, N_t - 1$):

$$RE_t = \frac{\sqrt{\sum_{k=0}^{N-1} \sum_{l=0}^{N_t-1} \left| \hat{u}^{k,l} - u \left(n_{k+1}, t_0 + l \frac{t_T/2 - t_0}{N_t - 1} \right) \right|^2}}{\sqrt{\sum_{k=0}^{N-1} \sum_{l=0}^{N_t-1} \left| u \left(n_{k+1}, t_0 + l \frac{t_T/2 - t_0}{N_t - 1} \right) \right|^2}}, \quad (13)$$

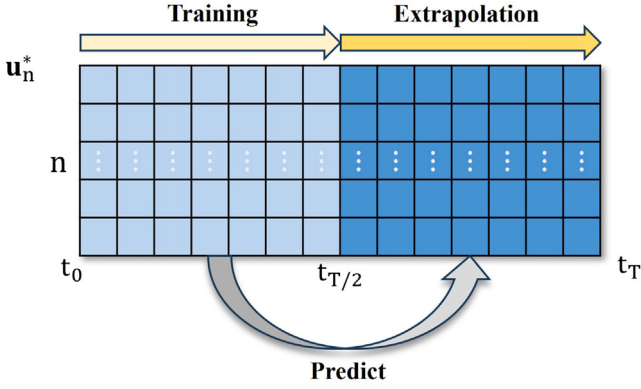


Fig. 4. The schematic diagram of the internal region training and extrapolation prediction.

the second stage involves evaluating the model's extrapolation generalization ability through the relative $\mathbb{L}_2$ error of $N \times N'_t$ ($N'_t = N_t$) data points on the given computational grid $\Omega \times [t_0, t_{T/2} + i \frac{t_T - t_{T/2}}{N'_t - 1}]$ ($i = 1, 2, \dots, N'_t - 1$):

$$RE_e = \frac{\sqrt{\sum_{k=0}^{N-1} \sum_{l=0}^{N'_t-1} \left| \hat{u}^{k, N_t+l-1} - u \left(n_{k+1}, t_1 + l \frac{t_T - t_{T/2}}{N'_t - 1} \right) \right|^2}}{\sqrt{\sum_{k=0}^{N-1} \sum_{l=0}^{N'_t-1} \left| u \left(n_{k+1}, t_1 + l \frac{t_T - t_{T/2}}{N'_t - 1} \right) \right|^2}}, \quad (14)$$

where $\hat{u}$ and $u(\cdot)$ stand for predictive value and true value respectively. Fig. 4 illustrates the schematic diagram of the internal region training and extrapolation prediction. Fig. 4 clearly shows that the model trained through iterations within the time domain $t \in [t_0, t_{T/2}]$ can be used for extrapolation predictions in the time domain $t \in [t_{T/2}, t_T]$.

2.5. Network settings

For the numerical results of any lattice equation presented below, we adopted the following network settings: in the encoder structure, the number of convolutional kernels in the three one-dimensional convolutional layers is 16, 64, 256, the number of channels in the first convolution operation is 2 (all discussed are coupled lattice equations), and the kernel size of the three convolution operations is 4, with a stride of 2 and zero-padding applied. A ConvLSTM layer is applied on the latent space with 128 nodes/cell/hidden states, with kernel sizes of 3, a stride of 1, and zero-padding applied. The decoder consists of a pixel shuffle operation with an upsampling factor of 8 and a convolution operation with a stride of 1, including two convolution kernels of size 5, with the number of channels matching the corresponding number of channels after the pixel shuffle operation. We chose the time steps in the interval $[t_0, t_{T/2}]$ for training and internal prediction, and in the interval $[t_{T/2}, t_T]$ for extrapolation prediction. The learning rate was set to 1×10^{-3} .

All code was executed in an environment with Python 3.9 and PyTorch 1.13. The numerical experiments were conducted on a professional computer equipped with GPU NVIDIA GeForce RTX 3090.

3. Numerical results without pseudo grid training

In this section, our aim is to obtain high-precision numerical results for integer grid points within a given domain and reliable extrapolated results beyond their time domain. To achieve this, we consider the training method with initial condition data from $n \in \mathbb{Z}$.

3.1. One-soliton solution of Toda lattice

The Toda equation describes a one-dimensional chain of particles, where each particle with a mass of 1 interacts with its nearest neighbors through an exponential potential. We consider the initial boundary value problem for one-dimensional Toda lattice with coupled form [5]:

$$\begin{cases} \dot{u}_n = u_n(v_n - v_{n-1}), \\ \dot{v}_n = u_{n+1} - u_n, \quad n \in \Omega, \quad t \in [t_0, t_T], \\ u_n(t_0) = u(n, t_0), \quad v_n(t_0) = v(n, t_0), \\ u_{n_1} = u_{b_1}, \quad u_{n_N} = u_{b_2}, \\ v_{n_1} = v_{b_1}, \quad v_{n_N} = v_{b_2}. \end{cases} \quad (15)$$

This model finds applications in describing the dynamical properties of various systems such as LC transmission lines [5], biopolymers (e.g., DNA chains [72]), excitations in anharmonic lattices, and lattices of optical solitons in fibers [73]. Besides, in the continuous limit (3), Eq. (15) approximates the KdV equation. Therefore, the discrete solutions of Eq. (15) can be used to indirectly explore the dynamical behavior of the solutions to the KdV equation. One soliton solution of Eq. (15) is [74]:

$$\begin{aligned} u_n &= \frac{\cosh[q_0(n - n_0(t))] \left\{ \sinh(q_0)e^{q_0(n - n_0(t) - 2)} + \cosh[q_0(n - n_0(t) - 1)] \right\} e^{-q_0}}{\cosh[q_0(n - n_0(t) - 1)]^2}, \\ v_n &= \frac{\sinh(q_0)^2}{\cosh[q_0(n - n_0(t) - 1)] \cosh[q_0(n - n_0(t))]}, \end{aligned} \quad (16)$$

where $n_0(t) = n_0 - \sinh(q_0)/q_0 t$. In this case, let $q_0 = 1/2$, $n_0 = 0$, $\Omega = [-12, 13]$ is a finite one-dimension lattice and $[t_0, t_1] = [-0.5, 0.5]$. Substituting these conditions into Eq. (16), we can obtain the initial value for Eq. (15). In addition, we default the boundary values of the potential functions u_n and v_n (in the constant background) to $u_{b_1} = u_{b_2} = 1$ and $v_{b_1} = v_{b_2} = 0$. A time step size of $\delta_t = 0.005$ is chosen in the time domain to account for the approximation of derivatives in the lattice equation and the implicit time progression in ConvLSTM. Therefore, the region we use as training is $R_{in}(R_{in} = [n, t], n \in [-11, 12], t \in [-0.5, 0])$ and the training is conducted with 100 time steps per epoch in the time interval $t \in [-0.5, 0]$. Additionally, predictions are made in region $R_{out}(R_{out} = [n, t], n \in [-11, 12], t \in [0, 0.5])$ to assess the extrapolation capability of the trained model.

After training for 2000 epochs, PhyCRNet achieves a significant reduction in the loss function, reaching $2.320E - 08$ after 1003 s. Subsequently, we conduct the predictions in the time domain $t \in [-0.5, 0]$ and obtained excellent numerical results, in which the $\mathbb{L}_2$ error of component u_n reaches $2.0567724E - 05$, and that of component v_n reaches $2.2417906E - 04$. Moreover, extrapolation predictions in the time domain $t \in [0, 0.5]$ also yield good numerical results, with the $\mathbb{L}_2$ error for component u_n at $1.0692835E - 03$, and for component v_n at $1.3986321E - 02$.

Figs. 5 and 6 respectively present the density plots, cross-sectional plots (temporal evolution plots), and error dynamics plots of one-soliton solutions for Toda lattice based on the PhyCRNet method. The density plot in (a) illustrates the evolution of the line soliton along the negative direction of the n -axis, which is clearly evident in the figure. Additionally, the consistency between the predicted and exact solutions in the cross-sectional plots (b1)–(b2) indicates that the model has achieved good performance in learning the one-soliton solution within the time domain $t \in [-0.5, 0]$, while (b3) demonstrates successful extrapolation of the one-soliton solution within the time domain $t \in [0, 0.5]$. From (c), it can be observed that the model's generalization error mainly arises from the long-term evolution of the soliton, suggesting that the neural network's ability to transmit information at different time steps needs improvement, and this error partly originates from

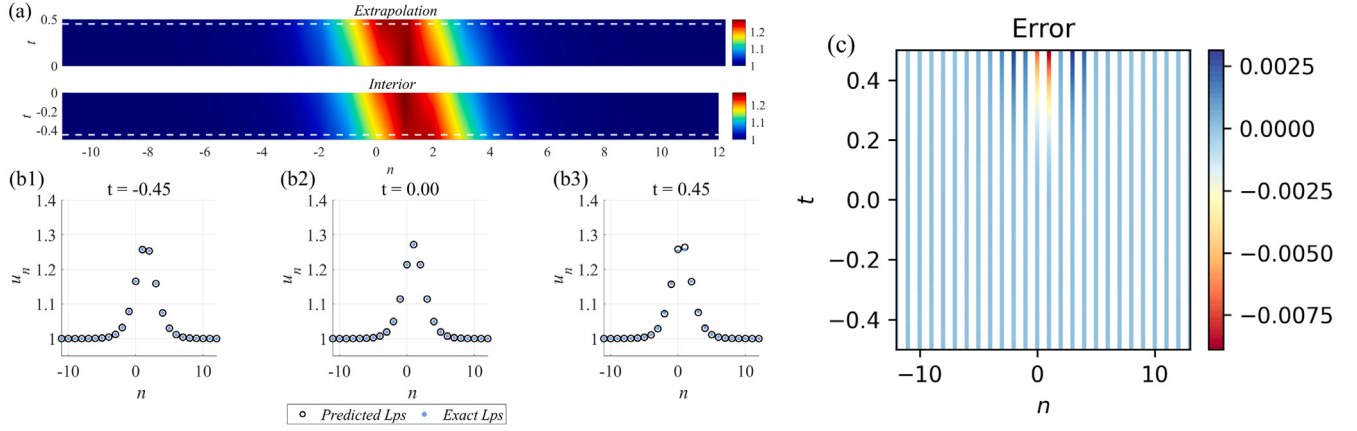


Fig. 5. The numerical one-soliton solution of the Toda lattice for the component u_n is presented. (a) illustrates the density plot of component u_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-0.5, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 0.5]$. The white dashed lines indicate the selected moments at $t = -0.45$ and $t = 0.45$; (b1)–(b3) show the temporal evolution of one-soliton at three different moments, where “Lps” stands for lattice points. (c) presents the density plot of the error (predicted solution minus exact solution).

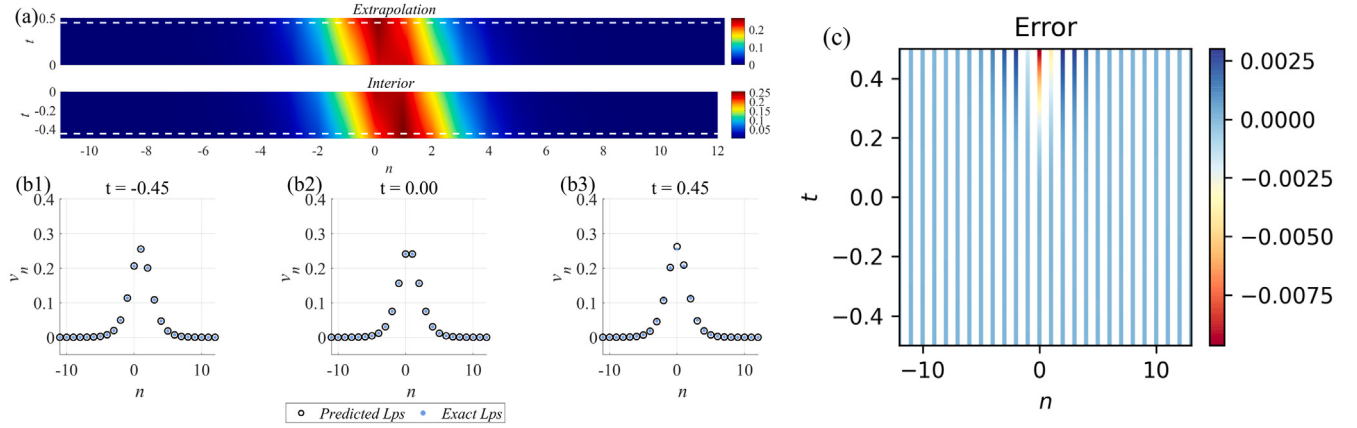


Fig. 6. The numerical one-soliton solution of the Toda lattice for the component v_n is presented. (a) illustrates the density plot of component v_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-0.5, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 0.5]$. The white dashed lines indicate the selected moments at $t = -0.45$ and $t = 0.45$; (b1)–(b3) show the temporal evolution of one-soliton at three different moments. (c) presents the density plot of the error (predicted solution minus exact solution).

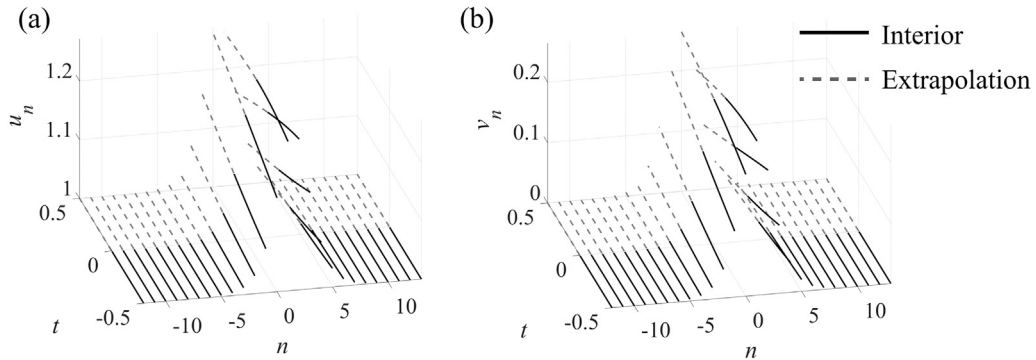


Fig. 7. Three-dimensional plots of the numerical solution of one-soliton for Toda lattice, where the solid black line indicates the prediction within the time domain $t \in [-0.5, 0]$, and the gray dashed line indicates the extrapolated prediction within the time domain $t \in [0, 0.5]$.

the differential results in the temporal direction. Furthermore, Fig. 7 shows the three-dimensional plots of the numerical solutions u_n and v_n , providing a clearer depiction of the line soliton’s dynamic behavior. It can be found in Table 1 the neural network has successfully learned the dynamic behavior of the single soliton solution on the Toda lattice, as supported by the evidence presented above.

Table 1
Errors of learning the one-soliton solution for Toda lattice.

	u_n	v_n	Total
Interior	2.0567724E-05	2.2417906E-04	2.7090538E-05
Extrapolation	1.0692835E-03	1.3986321E-02	1.5350957E-03

3.2. Dark-bright rational soliton solution of self-dual network equation

We investigate the self-dual network equation with the initial boundary value problem:

$$\begin{cases} \dot{I}_n = (1 + I_n^2)(V_{n-1} - V_n), \\ \dot{V}_n = (1 + V_n^2)(I_n - I_{n+1}), \quad n \in \Omega, \quad t \in [t_0, t_1], \\ I_n(t_0) = I(n, t_0), \quad V_n(t_0) = V(n, t_0), \\ I_{n_1} = I_{b_1}, \quad I_{n_N} = I_{b_2}, \\ V_{n_1} = V_{b_1}, \quad V_{n_N} = V_{b_2}. \end{cases} \quad (17)$$

The nonlinear self-dual network equation describes the propagation of electrical signals in a ladder-type electric circuit, where V_n and I_n represent the voltage and current [62]. Besides, in the continuous limit (6), Eq. (17) approximates the mKdV equation. Therefore, the discrete solutions of Eq. (17) can be used to indirectly explore the dynamical behavior of the solutions to the mKdV equation. Based on the Lax pair of Eq. (17) [6]

$$\begin{aligned} E\psi_n &= U_n\psi_n = \begin{pmatrix} \lambda + I_n V_n & -I_n + V_n/\lambda \\ I_n - \lambda V_n & 1/\lambda + I_n V_n \end{pmatrix} \psi_n, \\ \psi_n &= W_n\psi_n = \begin{pmatrix} \lambda - I_n V_{n-1} & -I_n + V_{n-1}/\lambda \\ I_n - \lambda V_{n-1} & 1/\lambda - I_n V_{n-1} \end{pmatrix} \psi_n, \end{aligned} \quad (18)$$

where E stands for shift operator, $\psi_n(t) = (\psi_{1,n}, \psi_{2,n})$ is the vector eigenfunction, λ is the eigenvalue parameter, the Darboux matrix $T_n(t; \lambda)$ can be defined as [75]:

$$T_n(t; \lambda) = \begin{pmatrix} \lambda^N + \sum_{j=0}^{N-1} A_n^{(j)} \lambda^j & \sum_{j=0}^{N-1} B_n^{(j)} \lambda^j \\ -\lambda^N \sum_{j=0}^{N-1} B_n^{(j)} \lambda^{-j} & 1 + \lambda^N \sum_{j=0}^{N-1} A_n^{(j)} \lambda^{-j} \end{pmatrix}, \quad (19)$$

where $A_n^{(j)}$ and $B_n^{(j)}$ can be determined by $2N$ linear system of equations obtained by Taylor expanding $\lim_{\epsilon \rightarrow 0} T(t; \lambda_1 + \epsilon)\psi_n(t; \lambda_1 + \epsilon)/\epsilon^{k_1} = 0$ ($k_1 = 0, 1, \dots, v_1$, e.g. $N = 1 + v_1$) at $\epsilon = 0$. Hence, the N -order rational soliton solution can be obtained by Darboux transformation of Eq. (17):

$$\hat{I}_n^{(N)} = \frac{I_n + B_n^{(N-1)}}{A_n^{(0)}}, \quad \hat{V}_n^{(N)} = B_{n+1}^{(0)} + V_n A_{n+1}^{(0)}. \quad (20)$$

Let $N = 1$, the unknown functions $A_n^{(0)} = \delta A_n^{(0)}/\delta_n$ and $B_n^{(0)} = \delta B_n^{(0)}/\delta_n$ are

$$\begin{aligned} \delta A_n^{(0)} &= \begin{vmatrix} -\lambda_1 \psi_{1,n}^{(0)} & \psi_{2,n}^{(0)} \\ -\psi_{2,n}^{(0)} & -\lambda_1 \psi_{1,n}^{(0)} \end{vmatrix}, \quad \delta B_n^{(0)} = \begin{vmatrix} \psi_{1,n}^{(0)} & -\lambda_1 \psi_{1,n}^{(0)} \\ \lambda_1 \psi_{2,n}^{(0)} & -\psi_{2,n}^{(0)} \end{vmatrix}, \\ \delta_n &= \begin{vmatrix} \psi_{1,n}^{(0)} & \psi_{2,n}^{(0)} \\ \lambda_1 \psi_{2,n}^{(0)} & -\lambda_1 \psi_{1,n}^{(0)} \end{vmatrix}, \end{aligned} \quad (21)$$

where the non-zero seed solutions $I_n = \alpha = \frac{11}{60}$, $V_n = 0$. The vector eigenfunction of Eq. (18) $\psi_n^{(0)} = (\psi_{1,n}^{(0)}, \psi_{2,n}^{(0)})$ can be determined by

$$\psi_n(t; \lambda, \epsilon)|_{\lambda=\lambda_1+\epsilon} = \begin{pmatrix} \psi_{1,n}^{(0)} \\ \psi_{2,n}^{(0)} \end{pmatrix} + \begin{pmatrix} \psi_{1,n}^{(1)} \\ \psi_{2,n}^{(1)} \end{pmatrix} \epsilon + \begin{pmatrix} \psi_{1,n}^{(2)} \\ \psi_{2,n}^{(2)} \end{pmatrix} \epsilon^2 + \dots, \quad \lambda_1 = \alpha + \sqrt{1 + \alpha^2}, \quad (22)$$

with

$$\begin{aligned} \psi_{1,n}^{(0)} &= \beta_n(t) \frac{671t + 660n + 3660}{60390}, \quad \psi_{2,n}^{(0)} = \beta_n(t) \frac{61t + 60n}{5490}, \\ \beta(n, t) &= \frac{61^{n+\frac{1}{2}} \sqrt{11}}{15^{n-\frac{1}{2}} 4^n} e^{\frac{61}{60}t}. \end{aligned} \quad (23)$$

Substituting Eq. (23) into (21) (20) and eliminating $\beta(n, t)$, we find that the mathematical expression for a rational soliton is entirely composed of rational polynomials. This is in contrast to traditional soliton expressions, which involve hyperbolic functions (in Eq. (16)). Of particular interest is whether the PhyCRNet method can effectively approximate this function and explore its dynamic behavior.

In this case, let $\Omega = [-31, 26]$ be a finite one-dimension lattice and $[t_0, t_1] = [-1, 1]$. The initial value for Eq. (15) can be obtained by

Table 2

Errors of learning the first-order rational solution for self-dual network equation.

	u_n	v_n	Total
Interior	8.877356E-04	9.527011E-04	9.160831E-04
Extrapolation	3.014642E-02	2.244632E-02	2.131525E-02

Eq. (20). In addition, we default the boundary values of the potential functions I_n and V_n (in the constant background) to $I_{b_1} = I_{b_2} = 0$ and $V_{b_1} = V_{b_2} = -\alpha$. A time step size of $\delta_t = 0.01$ is chosen in the time domain to account for the approximation of derivatives in the lattice equation and the implicit time progression in ConvLSTM. Therefore, the region we use as training is $R_{in}(R_{in} = [n, t], n \in [-30, 25], t \in [-1, 0])$ and the training is conducted with 100 time steps per epoch in the time interval $t \in [-1, 0]$. Additionally, predictions are made in region $R_{out}(R_{out} = [n, t], n \in [-30, 25], t \in [0, 1])$ to assess the extrapolation capability of the trained model. After training for 2000 epochs, PhyCRNet achieves a significant reduction in the loss function, reaching the order of $5.951E-07$ after 1138 s. Subsequently, we conduct the predictions in the time domain $t \in [-1, 0]$ and obtain excellent numerical results, in which the $\mathbb{L}_2$ error of component I_n reaches $8.877356E-04$, and that of component V_n reaches $9.527011E-04$. Moreover, extrapolation predictions in the time domain $t \in [0, 1]$ also yield good numerical results, with the $\mathbb{L}_2$ error for component I_n at $3.014642E-02$, and for component V_n at $2.244632E-02$.

Figs. 8 and 9 respectively present the density plots, cross-sectional plots (temporal evolution plots), and error dynamics plots of dark-bright rational soliton solutions for self-dual network equation based on the PhyCRNet method. The density plot in (a) illustrates the evolution of the rational soliton along the negative direction of the n -axis, which is clearly evident in the figure. Additionally, the consistency between the predicted and exact solutions in the cross-sectional plots (b1)–(b2) indicates that the model has achieved good performance in learning the rational soliton solution within the time domain $t \in [-1, 0]$, while (b3) demonstrates successful extrapolation of the rational soliton solution within the time domain $t \in [0, 1]$. In panel (c), errors arising from the propagation of rational soliton solutions over time can also be observed. Furthermore, Fig. 10 shows the three-dimensional plots of the numerical solutions I_n and V_n , providing a clearer depiction of the dark-bright rational soliton's dynamic behavior. It can be found in Table 2 the neural network has successfully learned the dynamic behavior of the rational soliton solution on the self-dual network equation, as supported by the evidence presented above.

4. Numerical results with pseudo grid training

After conducting several repeated experiments, it was observed that the model exhibited slight deficiencies in its extrapolation capabilities when dealing with soliton solutions characterized by higher wave speeds, narrower widths, or other complex features. Motivated by this observation, we aim for PG-PhyCRNet to effectively capture the evolution of solitons over time (i.e., its extrapolation capability in the temporal direction) in the training of soliton solutions with higher wave speeds or complexity. This may involve sacrificing a small portion of the training error within the given spatiotemporal region. In this section, we will continue to discuss different complex solitary waves using these two lattice models. The introduction of pseudo grid necessitates partitioning the spatial domain grid in Eqs. (15) and (17) as $\Omega = [n_1, n_1 + i \frac{n_N - n_1}{N_n - 1}]$, $i = 1, \dots, N_n - 1$. To investigate the effects of pseudo grid on the PG-PhyCRNet framework, we evaluate the performance differences between PhyCRNet with and without pseudo grid for the cases of two solitons in the Toda lattice and one- or two-solitons in the self-dual network equation.

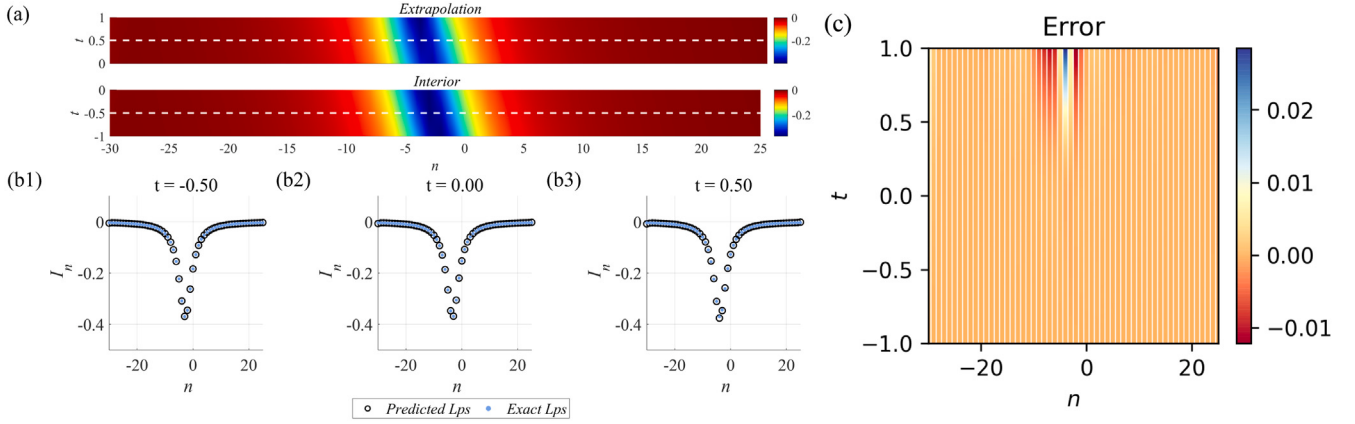


Fig. 8. The numerical dark rational soliton solution of the self-dual network equation for the component I_n is presented. (a) illustrates the density plot of component I_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-1, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 1]$. The white dashed lines indicate the selected moments at $t = -0.5$ and $t = 0.5$; (b1)–(b3) show the temporal evolution of rational soliton solution at three different moments. (c) presents the density plot of the error (predicted solution minus exact solution).

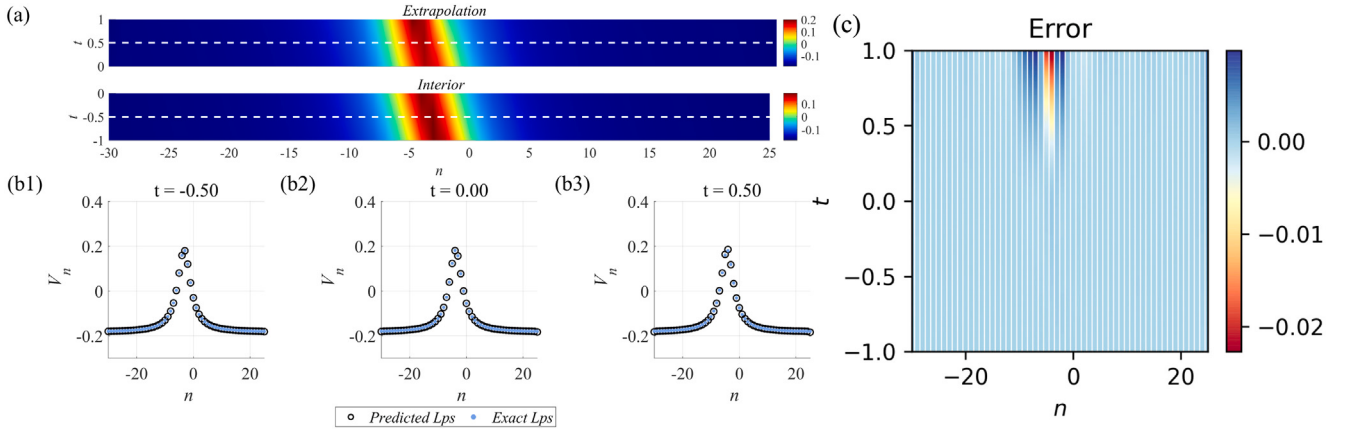


Fig. 9. The numerical bright rational soliton solution of the self-dual network equation for the component V_n is presented. (a) illustrates the density plot of component V_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-1, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 1]$. The white dashed lines indicate the selected moments at $t = -0.5$ and $t = 0.5$; (b1)–(b3) show the temporal evolution of rational soliton solution at three different moments. (c) presents the density plot of the error (predicted solution minus exact solution).

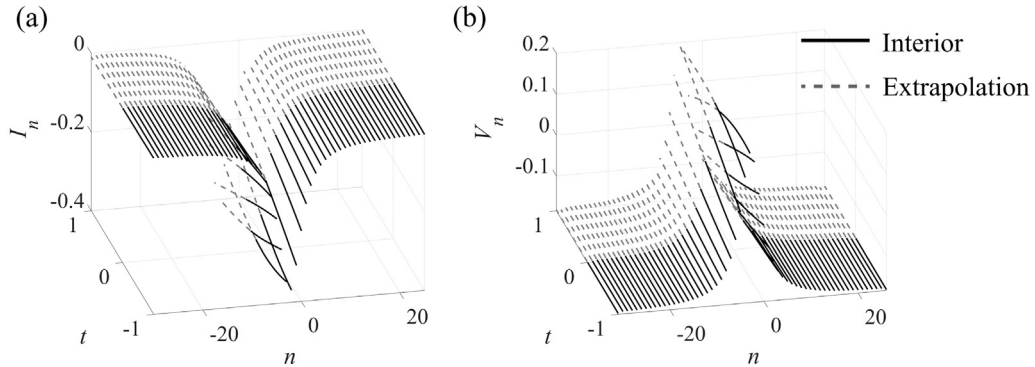


Fig. 10. Three-dimensional plots of the numerical solution of dark-bright rational soliton for self-dual network equation, where the solid black line indicates the prediction within the time domain $t \in [-1, 0]$, and the gray dashed line indicates the extrapolated prediction within the time domain $t \in [0, 1]$.

4.1. Two-soliton solution of Toda lattice under the non-zero background

In this section's numerical experiments, we will continue to discuss the initial-boundary value problem (15) for the one-dimensional Toda lattice. Based on the Darboux transformation in Ref. [76], we give the two-soliton solution of Toda lattice under the non-zero background:

$$\hat{u}_n = u_n - a_n + a_{n+1}, \quad \hat{v}_n = \frac{v_n + b_n}{1 + c_n}. \quad (24)$$

where

$$a_n = \frac{\lambda_2 \delta_{1,n} - \lambda_1 \delta_{2,n}}{\delta_{2,n} - \delta_{1,n}}, \quad b_n = \frac{\lambda_1 - \lambda_2}{\delta_{2,n} - \delta_{1,n}}, \quad c_n = \frac{(\lambda_2 - \lambda_1) \delta_{1,n} \delta_{2,n}}{(\lambda_1 - \lambda_2) \delta_{1,n} \delta_{2,n} + \delta_{2,n} - \delta_{1,n}},$$

$$\delta_{i,n} = \frac{r_i \tau_-(\lambda_i)^n e^{\rho_-(\lambda_i)t} - \tau_+(\lambda_i)^n e^{\rho_+(\lambda_i)t}}{\tau_+(\lambda_i)^n e^{\rho_+(\lambda_i)t} - r_i \tau_-(\lambda_i)^{n+1} e^{\rho_-(\lambda_i)t}},$$

$$\tau_{\pm}(\lambda_i) = 1 + \lambda_i \pm \frac{1}{2} \sqrt{\lambda_i^2 + 2\lambda_i - 3}, \quad \rho(\lambda_i)_{\pm} = \frac{\lambda_i \tau_{\pm} - 2}{2\tau_{\pm}} (i = 1, 2).$$

In this case, let $r_1 = -1$, $r_2 = 1$, $\lambda_1 = 2$, $\lambda_2 = 2.1$, the spatial domain $\Omega = [-12, 13]$ is divided into 225 segments ($n_1 = -12$, $n_N = 13$, $\delta_n = \frac{1}{9}$), while the time domain $[t_0, t_1] = [-1, 1]$ is divided into 200 segments (with a training time step of 100 and the remaining 100 segments used for extrapolation predictions). Substituting these conditions into Eq. (24), we can obtain the initial boundary value problem for Eq. (15). In addition, we default the boundary values of the potential functions u_n and v_n (in the constant background) to $u_{b_1} = u_{b_2} = 1$ and $v_{b_1} = v_{b_2} = -1$. A time step size of $\delta_t = 0.005$ is chosen in the time domain to account for the approximation of derivatives in the lattice and the implicit time progression in ConvLSTM. Therefore, the region we use as training is $R_{in}(R_{in} = [n, t], n \in [-12 + \delta_n, 13 - \delta_n], t \in [-1, 0])$ and the training is conducted with 100 time steps per epoch in the time interval $t \in [-1, 0]$. Additionally, the predictions are made in region $R_{out}(R_{out} = [n, t], n \in [-12 + \delta_n, 13 - \delta_n], t \in [0, 1])$ to assess the extrapolation capability of the trained model. After training for 1000 epochs, the loss reaches $2.446500E - 06$ after 688 s. Similarly, training and predictions are conducted in the time domain $t \in [-1, 0]$. The $\mathbb{L}_2$ error for component u_n reaches $1.462330E - 03$, while that for component v_n reaches $1.363670E - 03$. We further conduct the extrapolation predictions for the next 100 time steps. The $\mathbb{L}_2$ error of component u_n in the time domain $t \in [0, 1]$ reaches $2.730460E - 03$, and for component v_n reaches $2.687530E - 03$.

Figs. 11 and 12 respectively present the density plots, temporal evolution plots, and error dynamics plots of two-soliton solutions for Toda lattice based on the PG-PhyCRNet method. From (a), it can be observed that two nearly parallel solitons move towards the negative direction of space as time evolves. Figs. (b1)–(b3) respectively show the time evolution of lattice points with the assistance of pseudo grid (gray dashed lines), where (b1)–(b2) represent predictions within the training region, and (b3) represents extrapolation results outside the time domain. The consistency between the predicted and exact solutions indicates that the model has achieved good performance in learning the two-soliton solution. Furthermore, Fig. 13 shows the three-dimensional plots of the numerical solutions u_n and v_n . With an overall average generalization error of $1.431510E - 03$ for predictions within the time domain $t \in [-1, 0]$ and $2.716800E - 03$ for extrapolation outside the training range $t \in [0, 1]$, the neural network has successfully learned the dynamic behavior of two-soliton solution on the Toda lattice, as supported by the evidence presented above. The prediction accuracy for extrapolation is comparable to that within the region and does not exhibit significant differences.

Additionally, Table 3 compares the PG-PhyCRNet model's performance in learning the two-soliton solution of the Toda lattice with and without the assistance of pseudo grid. It is evident that while the PG-PhyCRNet model without grid assistance achieves higher precision in predicting the two-soliton solution within the training region, its accuracy in extrapolation beyond this region is unsatisfactory. Conversely, the PG-PhyCRNet model with grid assistance, although sacrificing some

precision in learning the two-soliton solution within the specified region, maintains a highly reliable extrapolation capability outside this region.

4.2. Multi-soliton solutions of self-dual network equation under the zero background

In this section's numerical experiments, we will continue to discuss the initial-boundary value problem (17). The N -soliton solution of Toda lattice under the zero background $I_n = V_n = 0$ also can be given by Eq. (20) and $A_n^{(0)} = \Delta A_n^{(0)} / \Delta_n$, $B_n^{(0)} = \Delta B_n^{(0)} / \Delta_n$, where

$$A_n = \begin{pmatrix} \lambda_1^{N-1} & \lambda_1^{N-2} & \cdots & 1 & \lambda_1^{N-1} \delta_{1,n} & \lambda_1^{N-2} \delta_{1,n} & \cdots & \delta_{1,n} \\ \lambda_2^{N-1} & \lambda_2^{N-2} & \cdots & 1 & \lambda_2^{N-1} \delta_{2,n} & \lambda_2^{N-2} \delta_{2,n} & \cdots & \delta_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_N^{N-1} & \lambda_N^{N-2} & \cdots & 1 & \lambda_N^{N-1} \delta_{N,n} & \lambda_N^{N-2} \delta_{N,n} & \cdots & \delta_{N,n} \\ \lambda_1 \delta_{1,n} & \lambda_1^2 \delta_{1,n} & \cdots & \lambda_1^N \delta_{1,n} & -\lambda_1 & -\lambda_1^2 & \cdots & -\lambda_1^N \\ \lambda_2 \delta_{2,n} & \lambda_2^2 \delta_{2,n} & \cdots & \lambda_2^N \delta_{2,n} & -\lambda_2 & -\lambda_2^2 & \cdots & -\lambda_2^N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_N \delta_{N,n} & \lambda_N^2 \delta_{N,n} & \cdots & \lambda_N^N \delta_{N,n} & -\lambda_N & -\lambda_N^2 & \cdots & -\lambda_N^N \end{pmatrix},$$

$$\delta_{i,n} = \lambda_i^{-2n} e^{-\frac{(i-1)(\lambda_i+1)t}{\lambda_i}}. \quad (25)$$

Here $\Delta A_n^{(0)}$, $\Delta B_n^{(0)}$, and $\Delta B_n^{(N-1)}$ can be obtained by replacing columns N , $2N$, and $N+1$ of A_n with the vector $(-\lambda_1^N, -\lambda_2^N, \dots, -\lambda_N^N, -\delta_{1,n}, -\delta_{2,n}, \dots, \delta_{N,n})^T$. Next, we primarily apply the PG-PhyCRNet method to verify the cases when $N = 1$ and $N = 2$, and utilize the pseudo-lattice training method to develop the one-, two-soliton model architecture of Eq. (17), demonstrating good extrapolation ability.

4.2.1. One-soliton solution

When $N = 1$ and $\lambda_1 = 2$, the one-soliton solution can be given by Eq. (20). The spatial domain $\Omega = [-12, 13]$ is divided into 325 segments, while the time domain $[t_0, t_1] = [-1, 1]$ is divided into 200 segments (with a training time step of 100 and the remaining 100 segments used for extrapolation predictions). After training for 1000 epochs, the loss reaches $6.717000E - 07$ after 698 s. Similarly, training and predictions are conducted in the time domain $t \in [-1, 0]$. The $\mathbb{L}_2$ error for component I_n reaches $2.672990E - 03$, while that for component V_n reaches $3.255200E - 03$. We further conduct the extrapolation predictions for the next 100 time steps. The $\mathbb{L}_2$ error of component I_n in the time domain $t \in [0, 1]$ reaches $5.417660E - 03$, and for component V_n reaches $6.182490E - 03$.

Figs. 14 and 15 present the density plots, temporal evolution plots, and error dynamics plots of the one-soliton solutions for Eq. (17) based on the PG-PhyCRNet method with pseudo grid. In (a), a line soliton can be seen moving in the negative spatial direction over time. Figures (b1)–(b3) depict the temporal evolution of lattice points with the aid of pseudo grid (gray dashed lines); (b1) and (b2) show predictions within the training region, while (b3) shows extrapolation results outside the time domain. The close match between predicted and exact solutions demonstrates that the model performs well in learning the one-soliton solution. Additionally, Fig. 16 provides three-dimensional plots of the numerical solutions I_n and V_n . With an overall average generalization error of $1.431510E - 03$ for predictions within the time domain $t \in [-1, 0]$ and $2.716800E - 03$ for extrapolation outside the training range $t \in [0, 1]$, the neural network successfully captures the dynamic behavior of the one-soliton solution. The prediction accuracy for extrapolation is comparable to that within the region and does not exhibit significant differences.

Table 4 compares the PG-PhyCRNet model's performance in learning the one-soliton solution of Eq. (17) with and without the assistance of pseudo grid. Similar to the numerical results in the previous subsection, the PG-PhyCRNet method assisted by pseudo grid performs significantly better in capturing the lattice point features outside the time domain when training the one-soliton of Eq. (17). It is observed that the one-soliton with a narrower wave width is more effectively

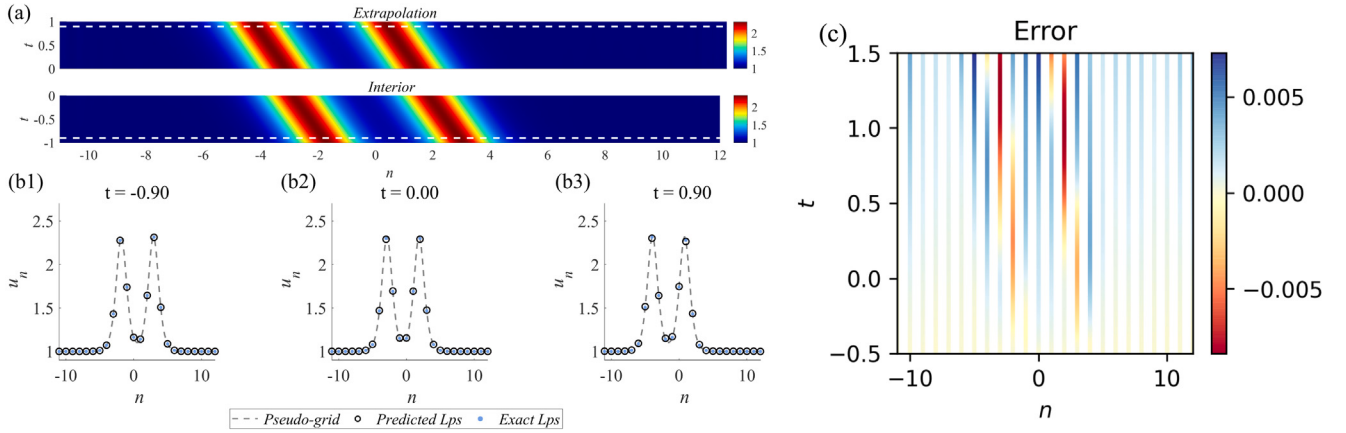


Fig. 11. The numerical two-soliton solution of Toda lattice for the component u_n is presented. (a) illustrates the density plot of component u_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-1, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 1]$. The white dashed lines indicate the selected moments at $t = -0.9$ and $t = 0.9$; (b1)–(b3) show the temporal evolution of two-soliton solution at three different moments. (c) presents the density plot of the lattice error (predicted solution minus exact solution).

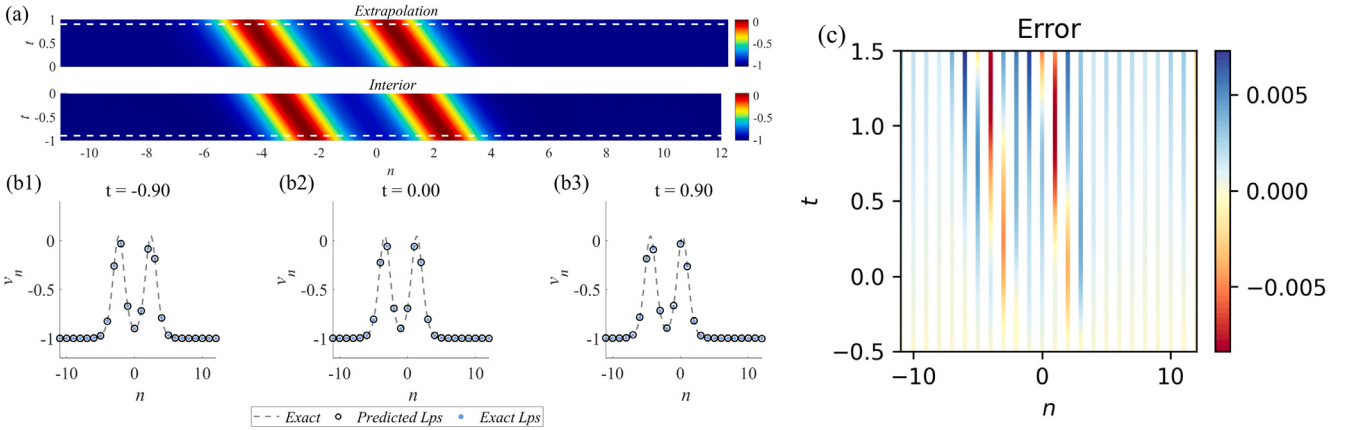


Fig. 12. The numerical two-soliton solution of Toda lattice for the component v_n is presented. (a) illustrates the density plot of component v_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-1, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 1]$. The white dashed lines indicate the selected moments at $t = -0.9$ and $t = 0.9$; (b1)–(b3) show the temporal evolution of two-soliton solution at three different moments. (c) presents the density plot of the lattice error (predicted solution minus exact solution).

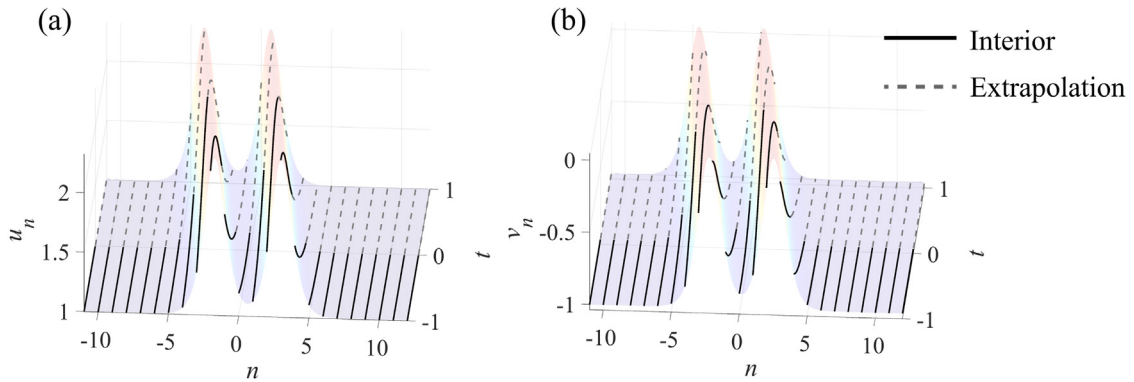


Fig. 13. Three-dimensional plots of the numerical solution of two-soliton for Toda lattice, where the solid black line indicates the prediction within the time domain $t \in [-1, 0]$, and the gray dashed line indicates the extrapolated prediction within the time domain $t \in [0, 1]$. Colored surfaces represent predicted solutions with pseudo grid.

Table 3

Errors of two-soliton solution for Toda lattice with or without the assistance of pseudo grid training (1000 epoch).

	Without pseudo grid (670 s for training time)			With pseudo grid (688 s for training time)		
	u_n	v_n	Total	u_n	v_n	Total
Interior	1.287950E-03	1.509930E-03	1.362890E-03	1.462330E-03	1.363670E-03	1.431510E-03
Extrapolation	9.076977E-02	1.046573E-01	9.543207E-02	2.730460E-03	2.687530E-03	2.716800E-03

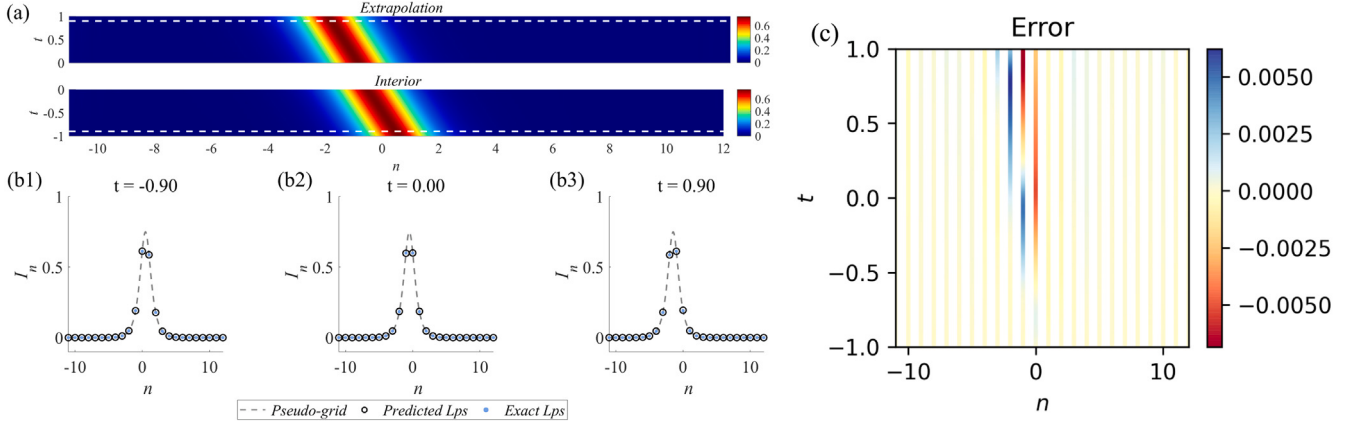


Fig. 14. The numerical one-soliton solution of self-dual network equation for the component I_n is presented. (a) illustrates the density plot of component I_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-1, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 1]$. The white dashed lines indicate the selected moments at $t = -0.9$ and $t = 0.9$; (b1)–(b3) show the temporal evolution of one-soliton solution at three different moments. (c) presents the density plot of the lattice error (predicted solution minus exact solution).

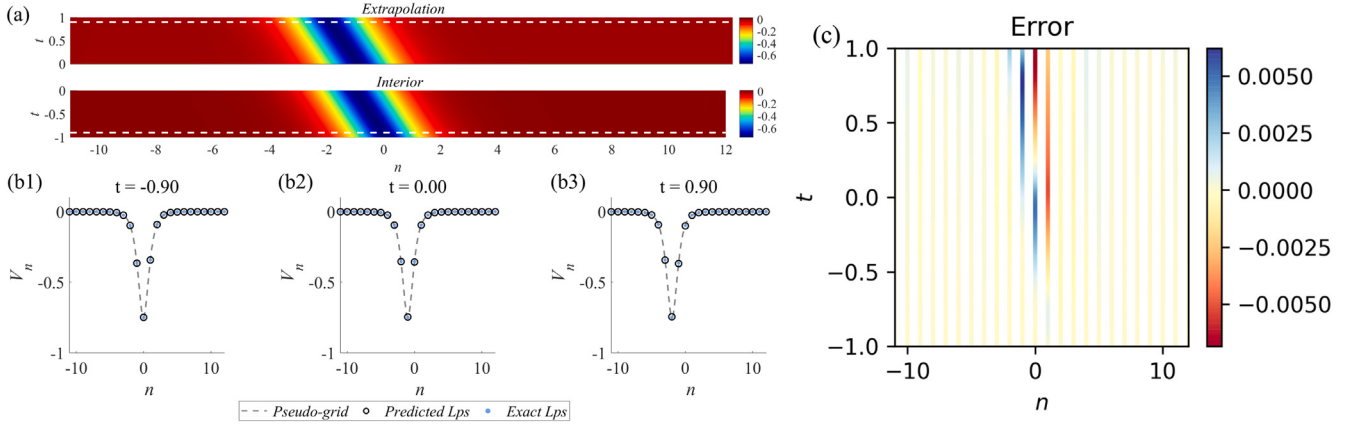


Fig. 15. The numerical one-soliton solution of self-dual network equation for the component V_n is presented. (a) illustrates the density plot of component V_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-1, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 1]$. The white dashed lines indicate the selected moments at $t = -0.9$ and $t = 0.9$; (b1)–(b3) show the temporal evolution of one-soliton solution at three different moments. (c) presents the density plot of the lattice error (predicted solution minus exact solution).

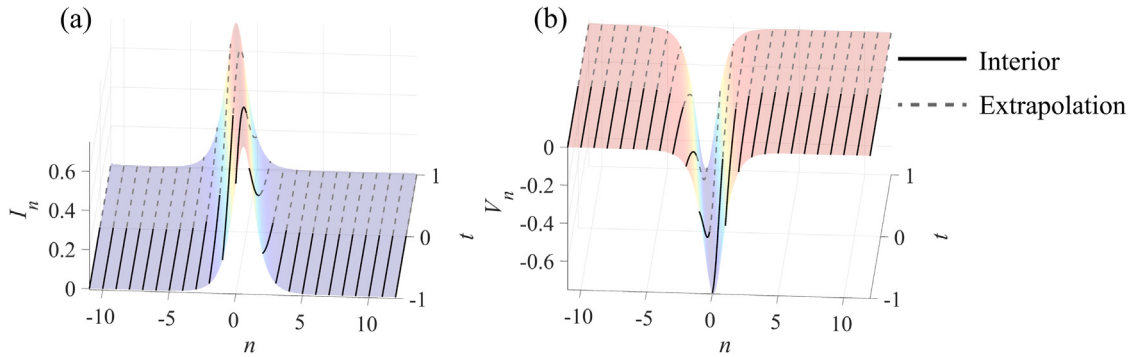


Fig. 16. Three-dimensional plots of the numerical solution of one-soliton for self-dual network equation, where the solid black line indicates the prediction within the time domain $t \in [-1, 0]$, and the gray dashed line indicates the extrapolated prediction within the time domain $t \in [0, 1]$. Colored surfaces represent predicted solutions with pseudo grid.

learned using the PG-PhyCRNet method with pseudo grid assistance. In contrast, the one-soliton solution trained using PhyCRNet without pseudo grid assistance is poorly predicted outside the time domain.

4.2.2. Two-soliton solution

When $N = 2$, $\lambda_1 = 1/5$ and $\lambda_2 = 6$, the two-soliton solution also can be given by Eq. (20). The spatial domain $\Omega = [-12, 13]$ is divided into 425 segments, while the time domain $[t_0, t_1] = [-0.5, 0.5]$

is divided into 200 segments (with a training time step of 100 and the remaining 100 segments used for extrapolation predictions). After training for 1000 epochs, the loss reaches $1.678600E-06$ after 702 s. Similarly, training and predictions are conducted in the time domain $t \in [-0.5, 0]$. The $\mathbb{L}_2$ error for component I_n reaches $1.162890E-03$, while that for component V_n reaches $7.419600E-04$. We further conduct the extrapolation predictions for the next 100 time steps. The $\mathbb{L}_2$ error

Table 4

Errors of one-soliton solution of self-dual network equation with or without the assistance of pseudo grid training (1000 epoch).

	Without pseudo grid (681 s for training time)			With pseudo grid (698 s for training time)		
	u_n	v_n	Total	u_n	v_n	Total
Interior	2.973470E-03	2.349200E-03	2.679000E-03	2.672990E-03	3.255200E-03	2.978890E-03
Extrapolation	2.519825E-01	2.248192E-01	2.387621E-01	5.417660E-03	6.182490E-03	5.813380E-03

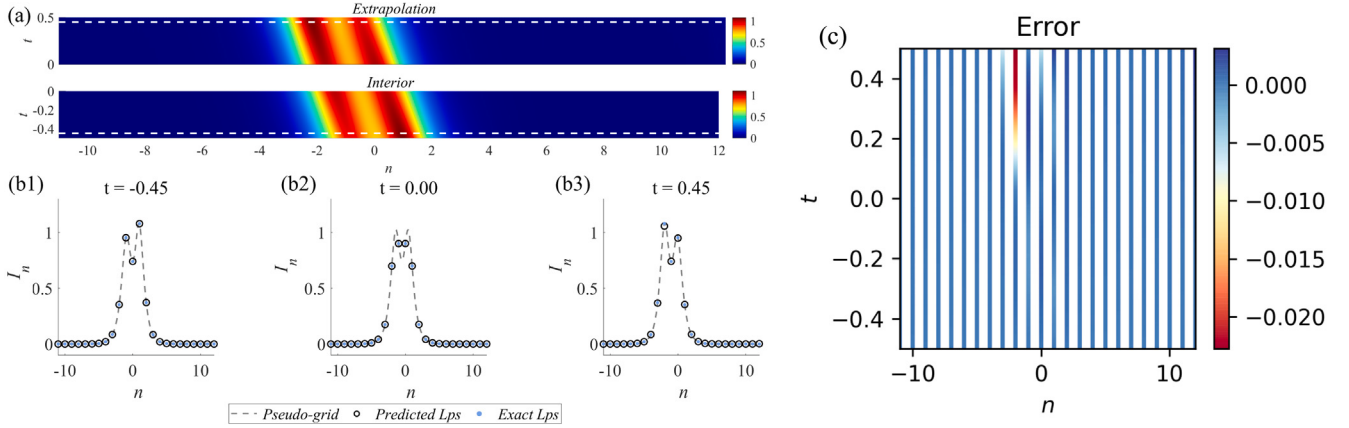


Fig. 17. The numerical two-soliton solution of self-dual network equation for the component I_n is presented. (a) illustrates the density plot of component I_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-0.5, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 0.5]$. The white dashed lines indicate the selected moments at $t = -0.45$ and $t = 0.45$; (b1)–(b3) show the temporal evolution of two-soliton solution at three different moments. (c) presents the density plot of the lattice error (predicted solution minus exact solution).

Table 5

Errors of two-soliton solution of self-dual network equation with or without the assistance of pseudo grid training (1000 epoch).

	Without pseudo grid (674 s for training time)			With pseudo grid (702 s for training time)		
	u_n	v_n	Total	u_n	v_n	Total
Interior	6.788400E-04	4.168900E-04	5.639000E-04	1.162890E-03	7.419600E-04	9.763600E-04
Extrapolation	6.503445E-02	1.116403E-01	9.125480E-02	9.618850E-03	1.556281E-02	1.292342E-02

of component I_n in the time domain $t \in [0, 0.5]$ reaches $9.618850E-03$, and for component V_n reaches $1.556281E-02$.

Figs. 17 and 18 present the density plots, temporal evolution plots, and error dynamics plots of the two-soliton solutions for Eq. (17) based on the PG-PhyCRNet method with pseudo grid. In (a), two soliton can be seen moving in the negative spatial direction over time. (b1)–(b3) depict the temporal evolution of lattice points with the aid of pseudo grid (gray dashed lines); (b1) and (b2) show predictions within the training region, while (b3) shows extrapolation results outside the time domain. The close match between predicted and exact solutions demonstrates that the model performs well in learning the two-soliton solution. Additionally, Fig. 19 provides three-dimensional plots of the numerical solutions I_n and V_n . With an overall average generalization error of $9.763600E-04$ for predictions within the time domain $t \in [-0.5, 0]$ and $1.292342E-02$ for extrapolation outside the training range $t \in [0, 0.5]$, the neural network successfully captures the dynamic behavior of the two-soliton solution. The prediction accuracy for extrapolation is comparable to that within the region and does not exhibit significant differences.

Table 5 compares the PG-PhyCRNet model's performance in learning the two-soliton solution of Eq. (17) with and without the assistance of pseudo grid. Similar to the numerical results in the previous subsection, the PG-PhyCRNet method assisted by pseudo grid performs significantly better in capturing the lattice point features outside the time domain when training the two-soliton of Eq. (17). It is observed that the one-soliton with a narrower wave width is more effectively learned using the PG-PhyCRNet method with pseudo grid assistance. In contrast, the two-soliton solution trained using PhyCRNet without pseudo grid assistance is poorly predicted outside the time domain.

5. Analysis and discussion

5.1. The effect of random initialization and spatial domain gridding

Since the weights of the neural network are initialized using Xavier initialization, the setting of the parameter seed in the code will affect the numerical results. Besides, the integer lattice points learn with boundary information, whereas the pseudo grid lack information during training, resulting in a high degree of freedom in their numerical solutions. To investigate the stability of integer lattice points in the presence of the pseudo grid in PG-PhyCRNet, numerical experiments with different randomly generated seed values and numbers of pseudo grid are conducted to observe their impact on the results (repeated experiments are shown in Tables 6–20 in the Appendix):

(1) The $\mathbb{L}_2$ generalization errors obtained under five different random seeds do not differ significantly, indicating that the degree of freedom of pseudo grid does not greatly impact the prediction and extrapolation of integer lattice points in multiple repeated experiments. Although some accuracy within the time domain may be lost or oscillations may appear in the background waves on pseudo grid, the PG-PhyCRNet method with the aid of pseudo grid still performs well in predicting solitary waves as mentioned in this paper.

(2) The prediction of different soliton solutions is affected by the density of pseudo grid. In Fig. 20, we show the effect of spatial gridding on the prediction of soliton solutions of the Toda lattice and self-dual network equations under different seeds. It can be observed that as the density of pseudo grid increases, the prediction error within the given spatiotemporal region increases until it stabilizes. Conversely, the extrapolation prediction error outside the time region decreases until it

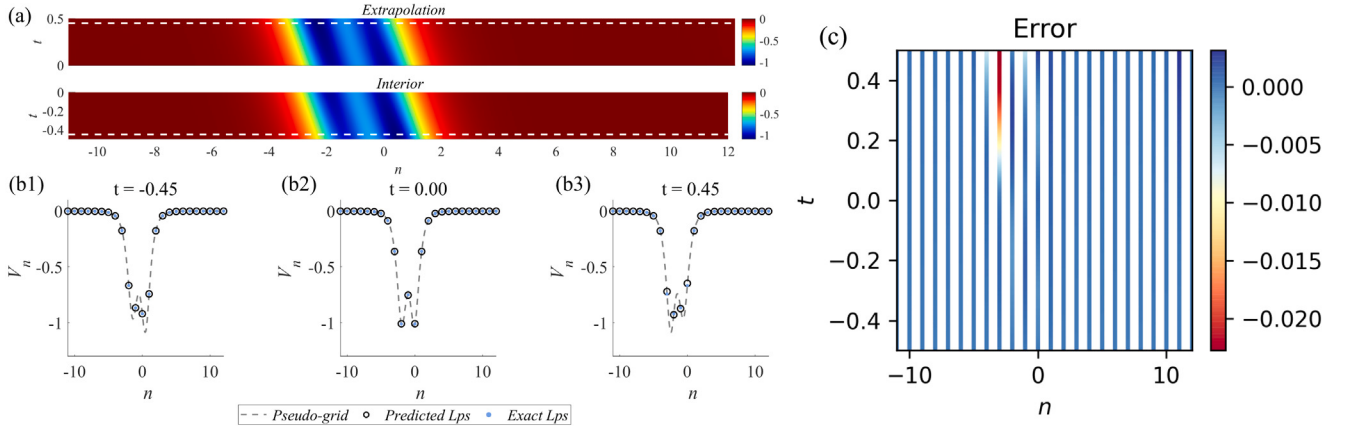


Fig. 18. The numerical two-soliton solution of self-dual network equation for the component V_n is presented. (a) illustrates the density plot of component V_n , with the lower sub-figure depicting the prediction within the time domain $t \in [-0.5, 0]$, and the upper sub-figure representing the extrapolated prediction within the time domain $t \in [0, 0.5]$. The white dashed lines indicate the selected moments at $t = -0.45$ and $t = 0.45$; (b1)–(b3) show the temporal evolution of two-soliton solution at three different moments. (c) presents the density plot of the lattice error (predicted solution minus exact solution).

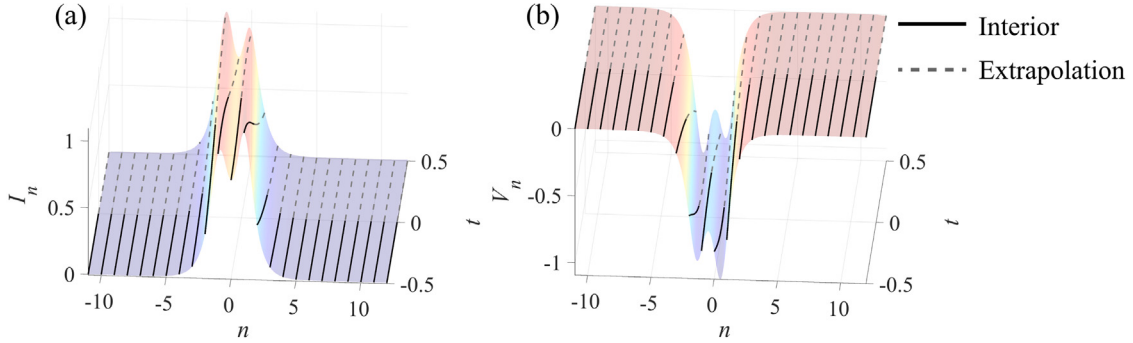


Fig. 19. Three-dimensional plots of the numerical solution of two-soliton for self-dual network equation, where the solid black line indicates the prediction within the time domain $t \in [-0.5, 0]$, and the gray dashed line indicates the extrapolated prediction within the time domain $t \in [0, 0.5]$. Colored surfaces represent predicted solutions with pseudo grid.

stabilizes. Therefore, it is necessary to choose an appropriate number of pseudo grid to participate in the model training.

5.2. Remark

The presence of excessive pseudo grid can lead to their weights being overly emphasized in the loss function, while the weights of integer lattice points are relatively diminished. Consequently, there is a certain loss when learning the integer lattice points within the spatiotemporal domain. One significant characteristic of CNNs is weight sharing, which imposes constraints on the weights of convolutional kernels by the integer lattice points during the training process. Therefore, the accuracy of learning the integer lattice points within the spatiotemporal domain does not suffer severely. However, the existence of pseudo grid makes it easier for the convolutional kernels to capture the dynamic behavioral characteristics of solitons. In other words, the PG-PhyCRNet method trained with pseudo grid learns a regression problem of curves rather than the dynamics of individual lattice points. Consequently, when the width of a soliton is small, too few trend characteristics of integer lattice points are captured at its steep position (where the slope of the function at a certain point is sufficiently large). This explains why the extrapolated performance of PG-PhyCRNet without the assistance of pseudo grid is poor or significantly worse (see Tables 6–20). Conversely, PG-PhyCRNet with the assistance of pseudo grid makes it easier for the convolutional kernels to capture the evolutionary trend of solitons, thereby enhancing the extrapolation capability of integer lattice points beyond the time domain.

6. Conclusions

In conventional discrete learning methods, most discussions focus on continuous nonlinear PDEs. The dynamics of the solutions to the corresponding continuous equations are indirectly explored by directly discretizing (using difference schemes) the continuous equations. Therefore, another discrete format of nonlinear integrable equations: integrable NLEs, are discussed via the discrete deep learning methods (PhyCRNet method) in this paper. These NLEs avoid the stability and convergence issues that arise from discretizing continuous equations with finite difference schemes in numerical calculations.

In integrable NLEs, the translation operator replaces the higher-order dispersion terms in continuous equations. Therefore, based on the ideas presented in Ref. [38], we propose the PG-PhyCRNet method specifically designed for solving integrable lattice equations. This method also cleverly addresses the difficulties that different waveforms of lattice solitons present to discrete learning training. The addition of pseudo grid enables the model to comprehensively learn various lattice soliton solutions:

- For lattice solitons with slower wave speeds and smoother waveforms, we train and predict directly on integer lattice points. In numerical experiments, we used a one-soliton of the Toda lattice and rational solitons of the self-dual network equation. The numerical results show that both the internal predictions and extrapolated predictions are excellent within the given space–time domain.
- For lattice solitons with faster wave speeds and steeper waveforms, we train and predict using both pseudo grid and integer

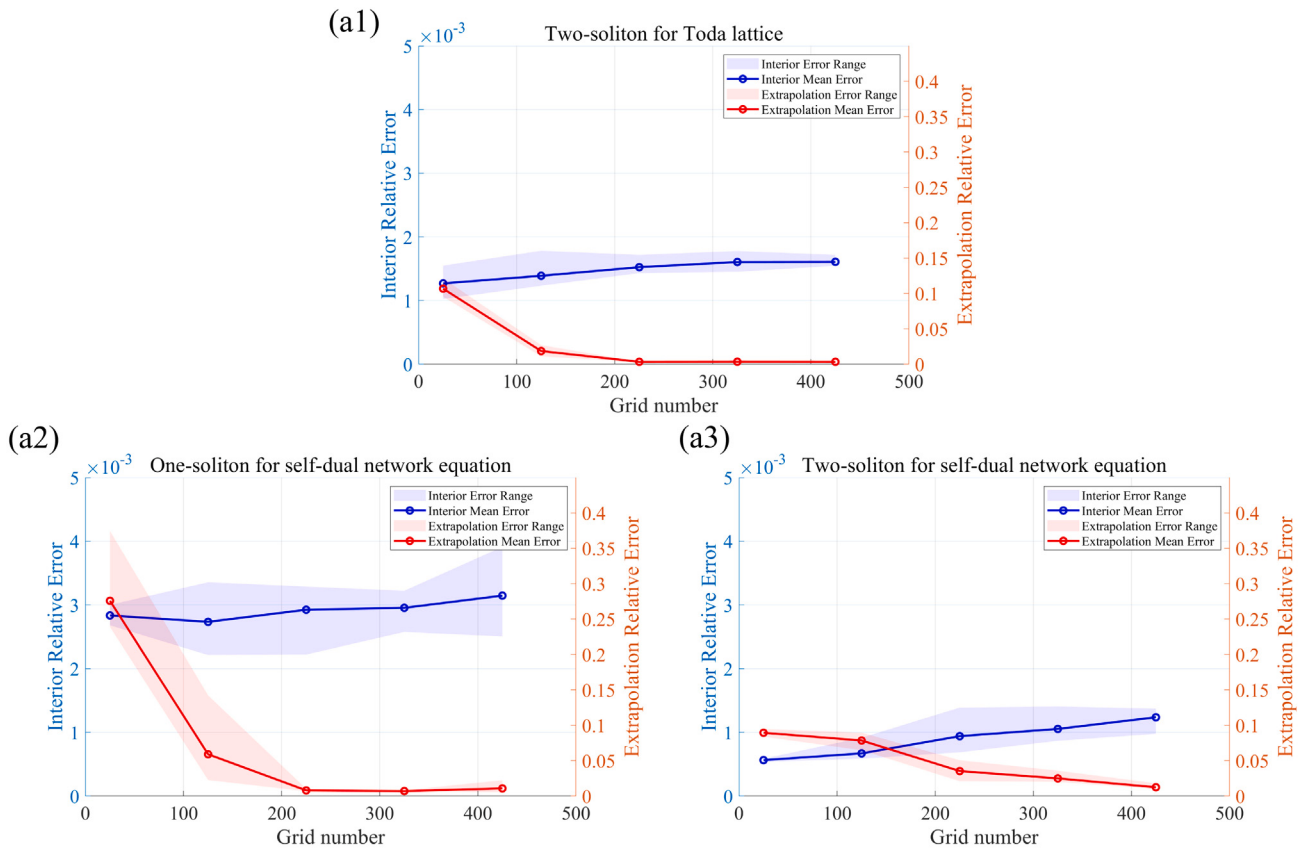


Fig. 20. Effect of spatial gridding on the prediction of soliton solutions of the Toda lattice and self-dual network equations under different seeds. The red region indicates the $\mathbb{L}_2$ error range predicted inside the given time domain, while the blue region indicates the $\mathbb{L}_2$ error range predicted outside the time domain.

lattice points. In numerical experiments, we used two-solitons of the Toda lattice and one- and two-solitons of the self-dual network equation. The numerical results indicate that the method of training with pseudo grid significantly outperforms the method without pseudo grid in terms of extrapolation capability.

The weight sharing in CNNs alleviates the errors' balance between pseudo grid and integer lattice points in PG-PhyCRNet training. Through the repeated experiments, PG-PhyCRNet with pseudo grid method sacrifices a small portion of training accuracy in the spatiotemporal domain compared to the version without pseudo grid. However, it exhibits excellent performance in capturing soliton trends and enhancing the model's extrapolation capability for solitons.

The PG-PhyCRNet method proposed for integrable NLEs is very necessary, as it provides a new perspective for solving continuous integrable systems using discrete deep learning methods. Of course, the numerical experiments in this paper only involve solving different types and structures of solitons. It is worth studying the generalization capability of this model for more complex localized wave solutions, such as rogue waves, breathers, and interaction solutions. Additionally, embedding certain properties of integrable discrete equations (such as discrete conservation laws, discrete Hamiltonian structures, continuous limits, and recursion operators) into the network to achieve different goals is something we aim to explore in the future.

CRediT authorship contribution statement

Zhe Lin: Methodology, Software, Writing – original draft, Writing – review & editing. **Yong Chen:** Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The project is supported by the National Natural Science Foundation of China [http://dx.doi.org/10.13039/501100001809] (No. 12175069 and No. 12235007), Science and Technology Commission of Shanghai Municipality, China (No. 21JC1402500 and No. 22DZ2229014), and Natural Science Foundation of Shanghai, China (No. 23ZR1418100).

Appendix A. Repeated experimental and relative $\mathbb{L}_2$ errors for training two-soliton solution for Toda lattice in the spatial domain $[-12, 13]$

See Tables 6–10.

Appendix B. Repeated experimental and relative $\mathbb{L}_2$ errors for training one-soliton solution for self-dual network equation in the spatial domain $[-12, 13]$

See Tables 11–15.

Table 6

Based on 25 grid numbers' training.

Seed	Interior			Extrapolation		
	u_n	v_n	Total	u_n	v_n	Total
100	1.004390E-03	1.066160E-03	1.024560E-03	9.199270E-02	1.062825E-01	9.679334E-02
1100	1.141840E-03	1.187920E-03	1.156780E-03	9.432476E-02	1.106409E-01	9.983428E-02
2100	1.287950E-03	1.509930E-03	1.362890E-03	9.076977E-02	1.046573E-01	9.543207E-02
3100	1.166190E-03	1.433770E-03	1.257990E-03	1.198982E-01	1.246799E-01	1.214482E-01
4100	1.620150E-03	1.387780E-03	1.549620E-03	1.182343E-01	1.264537E-01	1.209242E-01

Table 7

Based on 125 grid numbers' training.

Seed	Interior			Extrapolation		
	u_n	v_n	Total	u_n	v_n	Total
100	1.358250E-03	1.373920E-03	1.363290E-03	9.823230E-03	1.318727E-02	1.101165E-02
1100	1.290880E-03	1.225100E-03	1.270210E-03	2.419791E-02	3.235194E-02	2.707461E-02
2100	1.295460E-03	1.089890E-03	1.233440E-03	1.211099E-02	1.750363E-02	1.406270E-02
3100	1.346560E-03	1.188820E-03	1.298190E-03	1.712007E-02	2.405495E-02	1.960698E-02
4100	1.891910E-03	1.528060E-03	1.783620E-03	1.975810E-02	2.388229E-02	2.116490E-02

Table 8

Based on 225 grid numbers' training.

Seed	Interior			Extrapolation		
	u_n	v_n	Total	u_n	v_n	Total
100	1.454350E-03	1.575410E-03	1.494140E-03	3.424500E-03	3.625440E-03	3.490030E-03
1100	1.783300E-03	1.573400E-03	1.718950E-03	3.590040E-03	3.274400E-03	3.492180E-03
2100	1.462330E-03	1.363670E-03	1.431510E-03	2.730460E-03	2.687530E-03	2.716800E-03
3100	1.462900E-03	1.434220E-03	1.453790E-03	3.558970E-03	3.347580E-03	3.492750E-03
4100	1.597290E-03	1.370320E-03	1.528360E-03	3.498280E-03	3.283320E-03	3.430990E-03

Table 9

Based on 325 grid numbers' training.

Seed	Interior			Extrapolation		
	u_n	v_n	Total	u_n	v_n	Total
100	1.789990E-03	1.458460E-03	1.691030E-03	3.454720E-03	3.451360E-03	3.453650E-03
1100	1.762830E-03	1.286240E-03	1.625650E-03	3.959900E-03	2.963380E-03	3.670690E-03
2100	1.523710E-03	1.285470E-03	1.451770E-03	2.942450E-03	3.027140E-03	2.969800E-03
3100	1.479490E-03	1.489380E-03	1.482660E-03	2.669840E-03	3.111110E-03	2.818510E-03
4100	1.748590E-03	1.844100E-03	1.779700E-03	4.281690E-03	4.658000E-03	4.405560E-03

Table 10

Based on 425 grid numbers' training.

Seed	Interior			Extrapolation		
	u_n	v_n	Total	u_n	v_n	Total
100	1.645460E-03	1.670310E-03	1.653450E-03	3.941530E-03	3.485480E-03	3.801610E-03
1100	1.760540E-03	1.628470E-03	1.719400E-03	3.398780E-03	3.257960E-03	3.354380E-03
2100	1.559400E-03	1.554180E-03	1.557730E-03	3.227280E-03	3.319680E-03	3.257120E-03
3100	1.570980E-03	1.491500E-03	1.546000E-03	2.995290E-03	3.611780E-03	3.205410E-03
4100	1.588130E-03	1.524420E-03	1.568030E-03	2.973770E-03	3.200140E-03	3.048010E-03

Table 11

Based on 25 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	2.973470E-03	2.349200E-03	2.679000E-03	2.519825E-01	2.248192E-01	2.387621E-01
1100	3.114740E-03	2.687380E-03	2.908520E-03	3.546937E-01	3.933042E-01	3.745328E-01
2100	2.545430E-03	3.396040E-03	3.001810E-03	2.538650E-01	2.427684E-01	2.483683E-01
3100	2.603690E-03	3.033760E-03	2.827310E-03	2.831439E-01	2.339709E-01	2.596781E-01
4100	2.919450E-03	2.578170E-03	2.753790E-03	2.595958E-01	2.557138E-01	2.576585E-01

Table 12

Based on 125 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	I_n	V_n	Total
100	2.089710E-03	3.179500E-03	2.691360E-03	2.655297E-02	1.768221E-02	2.254981E-02
1100	2.173330E-03	2.301320E-03	2.238360E-03	6.974361E-02	3.462078E-02	5.502694E-02
2100	2.144920E-03	2.278520E-03	2.212860E-03	3.404159E-02	4.308837E-02	3.883774E-02
3100	2.269970E-03	3.865790E-03	3.171380E-03	1.751873E-01	9.917303E-02	1.422795E-01
4100	3.288780E-03	3.420710E-03	3.355510E-03	3.574693E-02	3.701973E-02	3.639008E-02

Table 13

Based on 225 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	2.028130E-03	3.893660E-03	3.106000E-03	9.274210E-03	8.104070E-03	8.707720E-03
1100	1.755860E-03	2.602480E-03	2.220680E-03	5.947070E-03	7.775850E-03	6.923810E-03
2100	2.720080E-03	3.534910E-03	3.154680E-03	6.701000E-03	7.475810E-03	7.099710E-03
3100	2.570200E-03	3.872700E-03	3.287810E-03	1.070728E-02	9.861580E-03	1.029233E-02
4100	2.356570E-03	3.270510E-03	2.851250E-03	6.671430E-03	6.627350E-03	6.649380E-03

Table 14

Based on 325 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	2.672990E-03	3.255200E-03	2.978890E-03	5.417660E-03	6.182490E-03	5.813380E-03
1100	2.551020E-03	3.415600E-03	3.015270E-03	6.496750E-03	8.152780E-03	7.372960E-03
2100	2.662000E-03	3.262530E-03	2.978000E-03	5.536020E-03	6.573370E-03	6.077840E-03
3100	2.680300E-03	3.684550E-03	3.222720E-03	8.051520E-03	8.241780E-03	8.147390E-03
4100	1.864270E-03	3.125680E-03	2.574600E-03	6.860780E-03	8.198070E-03	7.560300E-03

Table 15

Based on 425 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	4.203370E-03	3.599190E-03	3.912400E-03	1.030187E-02	9.803130E-03	1.005513E-02
1100	2.693260E-03	3.412580E-03	3.074700E-03	6.672860E-03	6.490600E-03	6.582190E-03
2100	2.395680E-03	3.745470E-03	3.145100E-03	1.060885E-02	8.064490E-03	9.420590E-03
3100	2.415120E-03	3.642350E-03	3.091390E-03	2.550555E-02	1.864582E-02	2.233421E-02
4100	2.034670E-03	2.896380E-03	2.503680E-03	5.263380E-03	5.644690E-03	5.457720E-03

Table 16

Based on 25 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	6.723700E-04	3.469900E-04	5.357300E-04	6.196126E-02	1.012896E-01	8.387203E-02
1100	6.788400E-04	4.168900E-04	5.639000E-04	6.503445E-02	1.116403E-01	9.125480E-02
2100	6.811500E-04	4.732600E-04	5.869600E-04	6.264318E-02	1.059869E-01	8.695842E-02
3100	6.901300E-04	3.579200E-04	5.504500E-04	6.329838E-02	1.111489E-01	9.033848E-02
4100	6.852900E-04	4.495200E-04	5.800600E-04	6.519826E-02	1.171786E-01	9.470403E-02

Table 17

Based on 125 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	6.849500E-04	5.057600E-04	6.024700E-04	5.959053E-02	1.000647E-01	8.226188E-02
1100	6.144300E-04	6.327200E-04	6.236000E-04	4.975587E-02	1.059173E-01	8.262451E-02
2100	7.169700E-04	1.080460E-03	9.160800E-04	3.486782E-02	9.738489E-02	7.301121E-02
3100	7.115200E-04	4.221800E-04	5.856700E-04	3.314541E-02	8.553515E-02	6.475365E-02
4100	6.964000E-04	5.303800E-04	6.193600E-04	6.024640E-02	1.116978E-01	8.962430E-02

Appendix C. Repeated experimental and relative $\mathbb{L}_2$ errors for training two-soliton solution for self-dual network equation in the spatial domain $[-12, 13]$

See [Tables 16–20](#).

Table 18
Based on 225 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	9.330600E-04	6.861400E-04	8.195200E-04	4.437688E-02	4.155745E-02	4.299681E-02
1100	6.980900E-04	6.734000E-04	6.859100E-04	1.297905E-02	2.655921E-02	2.087294E-02
2100	9.907600E-04	1.158130E-03	1.077310E-03	1.468222E-02	3.614822E-02	2.754275E-02
3100	1.015390E-03	1.683930E-03	1.388930E-03	4.295351E-02	2.244343E-02	3.431418E-02
4100	7.889300E-04	6.640200E-04	7.294500E-04	4.236627E-02	5.771308E-02	5.058945E-02

Table 19
Based on 325 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	8.840300E-04	1.253540E-03	1.083790E-03	1.548189E-02	2.359871E-02	1.993888E-02
1100	1.013320E-03	1.068150E-03	1.040970E-03	1.521999E-02	2.901112E-02	2.313511E-02
2100	6.687600E-04	1.028990E-03	8.669600E-04	1.875190E-02	2.473241E-02	2.193309E-02
3100	1.055200E-03	1.688590E-03	1.406540E-03	1.844988E-02	2.791708E-02	2.364030E-02
4100	8.523800E-04	8.927600E-04	8.727100E-04	2.094418E-02	4.615326E-02	3.578369E-02

Table 20
Based on 425 grid numbers' training.

Seed	Interior			Extrapolation		
	I_n	V_n	Total	T_n	V_n	Total
100	9.886800E-04	1.558600E-03	1.303840E-03	9.613700E-03	1.143157E-02	1.055762E-02
1100	1.162890E-03	7.419600E-04	9.763600E-04	9.618850E-03	1.556281E-02	1.292342E-02
2100	6.929300E-04	1.483680E-03	1.156170E-03	1.360377E-02	2.164958E-02	1.806173E-02
3100	1.106100E-03	1.597920E-03	1.373070E-03	9.200870E-03	1.103213E-02	1.015362E-02
4100	9.166300E-04	1.315060E-03	1.132580E-03	2.262608E-02	3.000274E-02	2.655471E-02

References

- [1] M. Raissi, P. Perdikaris, G. 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.
- [2] Z.Y. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, 2020, arXiv preprint arXiv:2010.08895.
- [3] L. Lu, P. Jin, G. Pang, Z. Zhang, G.E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nat. Mach. Intell.* 3 (3) (2021) 218–229.
- [4] A. Scott, *Encyclopedia of Nonlinear Science*, Routledge Press, London, 2004.
- [5] M. Toda, *Theory of Nonlinear Lattices*, Springer, Berlin, 1989.
- [6] M.J. Ablowitz, J.F. Ladik, Nonlinear differential-difference equations, *J. Math. Phys.* 16 (1975) 598–603.
- [7] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Netw.* 2 (5) (1989) 359–366.
- [8] Y. Bar-Sinai, S. Hoyer, J. Hickey, M. Brenner, Learning data-driven discretizations for partial differential equations, *Proc. Natl. Acad. Sci.* 116 (31) (2019) 15344–15349.
- [9] S. Wang, H. Wang, P. Perdikaris, Learning the solution operator of parametric partial differential equations with physics-informed DeepONets, *Sci. Adv.* 7 (40) (2021) eabi8605.
- [10] X. Jin, S. Cai, H. Li, G. Karniadakis, NSFnets (Navier-Stokes Flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations, *J. Comput. Phys.* 426 (2021) 109951.
- [11] Z. Mao, A. Jagtap, G. Karniadakis, Physics-informed neural networks for high-speed flows, *Comput. Methods Appl. Mech. Engrg.* 360 (2020) 112789.
- [12] E. Haghighat, M. Raissi, A. Moure, H. Gomez, R. Juanes, A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics, *Comput. Methods Appl. Mech. Engrg.* 379 (2021) 113741.
- [13] S. Cai, H. Li, F. Zheng, F. Kong, M. Dao, G. Karniadakis, S. Suresh, Artificial intelligence velocimetry and microaneurysm-on-a-chip for three-dimensional analysis of blood flow in physiology and disease, *Proc. Natl. Acad. Sci.* 118 (2021) e2100697118.
- [14] J. Li, Y. Chen, A physics-constrained deep residual network for solving the sine-gordon equation, *Commun. Theor. Phys.* 73 (1) (2020) 015001.
- [15] J.C. Pu, Y. Chen, Data-driven vector localized waves and parameters discovery for manakov system using deep learning approach, *Chaos Solitons Fractals* 160 (2022) 112182.
- [16] J.C. Pu, J. Li, Y. Chen, Solving localized wave solutions of the derivative nonlinear Schrödinger equation using an improved pinn method, *Nonlinear Dynam.* 105 (2021) 1723–1739.
- [17] J.C. Pu, J. Li, Y. Chen, Soliton, breather and rogue wave solutions for solving the nonlinear Schrödinger equation using a deep learning method with physical constraints, *Chin. Phys. B* 30 (2021) 060202.
- [18] H.J. Zhou, J.C. Pu, Y. Chen, Data-driven forward-inverse problems for the variable coefficients Hirota equation using deep learning method, *Nonlinear Dynam.* 111 (2023) 14667–14693.
- [19] H.J. Zhou, Parallel physics-informed neural networks method with regularization strategies for the forward-inverse problems of the variable coefficient modified KdV equation, *J. Syst. Sci. Complex.* 37 (2024) 511–544.
- [20] W.Q. Peng, Y. Chen, N-double poles solutions for nonlocal Hirota equation with nonzero boundary conditions using Riemann-Hilbert method and pinn algorithm, *Phys. D* 435 (2022) 133274.
- [21] Z.W. Miao, Y. Chen, Physics-informed neural networks method in high-dimensional integrable systems, *Modern Phys. Lett. B* 36 (01) (2022) 2150531.
- [22] Z.W. Miao, Y. Chen, VC-PINN: Variable coefficient physics-informed neural network for forward and inverse problems of PDEs with variable coefficient, *Phys. D* 456 (2023) 133945.
- [23] S.N. Lin, Y. Chen, The improved backward compatible physics-informed neural networks for reducing error accumulation and applications in data-driven higher-order rogue waves, *Chaos* 34 (2024) 033139.
- [24] S.N. Lin, Y. Chen, A two-stage physics-informed neural network method based on conserved quantities and applications in localized wave solutions, *J. Comput. Phys.* 457 (2022) 111053.
- [25] S.N. Lin, Y. Chen, Physics-informed neural network methods based on Miura transformations and discovery of new localized wave solutions, *Phys. D* 445 (2023) 133629.
- [26] J.C. Pu, Y. Chen, Lax pairs informed neural networks solving integrable systems, *J. Comput. Phys.* 510 (2024) 113090.
- [27] Z.J. Zhou, L. Wang, Z.Y. Yan, Deep neural networks learning forward and inverse problems of two-dimensional nonlinear wave equations with rational solitons, *Comput. Math. Appl.* 151 (2023) 164–171.
- [28] X.F. Zhou, C.Z. Qu, W.J. Yan, B. Li, Mastering the Cahn-Hilliard equation and Camassa-Holm equation with cell-average-based neural network method, *Nonlinear Dynam.* 111 (2023) 4823–4846.
- [29] M. Zhong, S. Gong, S.F. Tian, Z.Y. Yan, Data-driven rogue waves and parameters discovery in nearly integrable PT-symmetric Gross-Pitaevskii equations via PINNs deep learning, *Phys. D* 439 (2022) 133430.
- [30] J. Song, M. Zhong, G.E. Karniadakis, Z.Y. Yan, Two-stage initial-value iterative physics-informed neural networks for simulating solitary waves of nonlinear wave equations, *J. Comput. Phys.* 505 (2024) 112917.
- [31] Y.H. Zhu, N. Zabarar, Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification, *J. Comput. Phys.* 366 (2018) 415–447.

- [32] H. Gao, L.N. Sun, J.X. Wang, PhyGeoNet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain, *J. Comput. Phys.* 428 (2021) 110079.
- [33] N. Geneva, N. Zabarav, Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks, *J. Comput. Phys.* 403 (2020) 109056.
- [34] D. Kelshaw, L. Magri, Uncovering solutions from data corrupted by systematic errors: A physics-constrained convolutional neural network approach, 2023, arXiv preprint arXiv:2306.04600.
- [35] P.P. Shi, Z. Zeng, T.S. Liang, Physics-informed convnet: Learning physical field from a shallow neural network, *Commun. Nonlinear Sci. (2024)* 107911.
- [36] P. Ren, C.P. Rao, Y. Liu, Z.H. Ma, Q. Wang, J.X. Wang, H. Sun, PhysSR: Physics-informed deep super-resolution for spatiotemporal data, *J. Comput. Phys.* 492 (2023) 112438.
- [37] R.Y. Zhang, Y. Liu, H. Sun, Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling, *Eng. Struct.* 215 (2020) 110704.
- [38] P. Ren, C.P. Rao, Y. Liu, J.X. Wang, H. Sun, PhyCRNet: Physics-informed convolutional-recurrent network for solving spatiotemporal PDEs, *Comput. Methods Appl. Mech. Engrg.* 389 (2022) 114399.
- [39] E. Fermi, P. Pasta, S. Ulam, M. Tsingou, Studies of the Nonlinear Problems (No. LA-1940), Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 1989.
- [40] M. Toda, Wave propagation in anharmonic lattices, *J. Phys. Soc. Japan* 23 (1967) 501–506.
- [41] M. Toda, M. Wadati, A soliton and two solitons in an exponential lattice and related equations, *J. Phys. Soc. Japan* 34 (1973) 18–25.
- [42] M. Toda, Nonlinear lattice and soliton theory, *IEEE Trans. Circuit* 30 (1983) 542–554.
- [43] M.J. Ablowitz, J.F. Ladik, A nonlinear difference scheme and inverse scattering, *Stud. Appl. Math.* 55 (1976) 213–229.
- [44] A. Pickering, Z.N. Zhu, New integrable lattice hierarchies, *Phys. Lett. A* 349 (2002) 439–445.
- [45] D.J. Zhang, D.Y. Chen, The conservation laws of some discrete soliton systems, *Chaos Solitons Fractals* 14 (2002) 573–579.
- [46] K.M. Case, M. Kac, A discrete version of the inverse scattering problem, *J. Math. Phys.* 14 (5) (1973) 594–603.
- [47] R. Hirota, Nonlinear partial difference equations. I. A difference analogue of the Korteweg-de Vries equation, *J. Phys. Soc. Japan* 43 (4) (1977) 1424–1433.
- [48] R. Hirota, Nonlinear partial difference equations III; Discrete sine-Gordon equation, *J. Phys. Soc. Japan* 43 (6) (1977) 2079–2086.
- [49] M. Wadati, M. Watanabe, Conservation laws of a Volterra system and nonlinear self-dual network equation, *Progr. Theoret. Phys.* 57 (1977) 808–811.
- [50] M. Toda, Studies of a non-linear lattice, *Phys. Rep.* 18 (1975) 1–123.
- [51] M. Blaszk, K. Marciniak, R-matrix approach to lattice integrable systems, *J. Math. Phys.* 35 (1994) 4661–4682.
- [52] Y.B. Suris, O. Ragnisco, What is the relativistic Volterra lattice, *Comm. Math. Phys.* 200 (1999) 445–485.
- [53] H. Nagashima, Y. Amagishi, Experiment on the Toda lattice using nonlinear transmission lines, *J. Phys. Soc. Japan* 45 (1978) 680–688.
- [54] H. Spohn, Collision rate ansatz for the classical Toda lattice, *Phys. Rev. E* 101 (2020) 060103.
- [55] M.N. Sun, S.F. Deng, D.Y. Chen, The Bäcklund transformation and novel solutions for the Toda lattice, *Chaos Solitons Fractals* 23 (2005) 1169–1175.
- [56] Y.T. Wu, X.B. Hu, A new integrable differential-difference system and its explicit solutions, *J. Phys. A* 32 (1999) 1515–1521.
- [57] X.B. Hu, Y.T. Wu, X.G. Geng, Hirota bilinear approach to a new integrable differential-difference system, *J. Math. Phys.* 40 (1999) 2001–2010.
- [58] E.G. Fan, H.H. Dai, A differential-difference hierarchy associated with relativistic Toda and Volterra hierarchies, *Phys. Lett. A* 372 (2008) 4578–4585.
- [59] P. Liu, J. M., S.Y. Lou, A discrete Lax-integrable coupled system related to coupled KdV and coupled mKdV equations, *Chin. Phys. Lett.* 24 (2007) 2717.
- [60] D.J. Korteweg, G. de Vries, On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves, *Phil. Mag.* 39 (1895) 422.
- [61] H. Washimi, T. Taniuti, Propagation of ion-acoustic solitary waves of small amplitude, *Phys. Rev. Lett.* 17 (1966) 996.
- [62] R. Hirota, Exact N-soliton solution of nonlinear lumped self-dual network equations, *J. Phys. Soc. Japan* 35 (1) (1973) 289–294.
- [63] R.M. Miura, Korteweg-de Vries equation and generalizations. I. A remarkable explicit nonlinear transformation, *J. Math. Phys.* 9 (1968) 1202–1204.
- [64] M. Wadati, The modified Korteweg-de Vries equation, *J. Phys. Soc. Japan* 34 (1973) 1289–1296.
- [65] A.H. Khater, O.H. El-Kalaawy, D.K. Callebaut, Bäcklund transformations and exact solutions for Alfvén solitons in a relativistic electronpositron plasma, *Phys. Scr.* 58 (1998) 545–548.
- [66] S. Saqlain, W. Zhu, E.G. Charalampidis, P.G. Kevrekidis, Discovering governing equations in discrete systems using pinns, *Commun. Nonlinear Sci.* 126 (2023) 107498.
- [67] W. Zhu, W. Khademi, E.G. Charalampidis, P.G. Kevrekidis, Neural networks enforcing physical symmetries in nonlinear dynamical lattices: The case example of the Ablowitz-Ladik model, *Phys. D* 434 (2022) 133264.
- [68] R. Balestrieri, J. Pesenti, Y. LeCun, Learning in high dimension always amounts to extrapolation, 2021, arXiv preprint arXiv:2110.09485.
- [69] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [70] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W. c. Woo, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, *Adv. Neural Inf. Process. Syst.* 28 (2015) 802–810.
- [71] W. Shi, J. Caballero, F. Huszár, J. Totz, A. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.
- [72] V. Muto, P.S. Lomdahl, P.L. Christiansen, Two-dimensional discrete model for DNA dynamics: longitudinal wave propagation and denaturation, *Phys. Rev. A* 42 (12) (1990) 7452.
- [73] V.S. Gerdjikov, D.J. Kaup, I.M. Uzunov, E.G. Evstatiev, Asymptotic behavior of N-soliton trains of the nonlinear Schrödinger equation, *Phys. Rev. Lett.* 77 (19) (1996) 3943.
- [74] J. Garnier, F.K. Abdullaev, Soliton dynamics in a random Toda chain, *Phys. Rev. E* 67 (2) (2003) 026609.
- [75] X.Y. Wen, Z.Y. Yan, G.Q. Zhang, Nonlinear self-dual network equations: modulation instability, interactions of higher-order discrete vector rational solitons and dynamical behaviours, *Proc. R. Soc. A* 476 (2020) 20200512.
- [76] X.Y. Wen, N-fold Darboux transformation and soliton solutions for Toda lattice equation, *Rep. Math. Phys.* 68 (2) (2011) 211–223.