

Data-driven parametric soliton-rogon state transitions for nonlinear wave equations using deep learning with Fourier neural operator

Ming Zhong^{1,2}, Zhenya Yan^{1,2,*} and Shou-Fu Tian³

¹KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

²School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

³School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China

E-mail: zyyan@mmrc.iss.ac.cn

Received 28 July 2022, revised 27 October 2022

Accepted for publication 14 December 2022

Published 6 February 2023



CrossMark

Abstract

In this paper, we develop the deep learning-based Fourier neural operator (FNO) approach to find parametric mappings, which are used to approximately display abundant wave structures in the nonlinear Schrödinger (NLS) equation, Hirota equation, and NLS equation with the generalized \mathcal{PT} -symmetric Scarf-II potentials. Specifically, we analyze the state transitions of different types of solitons (e.g. bright solitons, breathers, peakons, rogons, and periodic waves) appearing in these complex nonlinear wave equations. By checking the absolute errors between the predicted solutions and exact solutions, we can find that the FNO with the GeLu activation function can perform well in all cases even though these solution parameters have strong influences on the wave structures. Moreover, we find that the approximation errors via the physics-informed neural networks (PINNs) are similar in magnitude to those of the FNO. However, the FNO can learn the entire family of solutions under a given distribution every time, while the PINNs can only learn some specific solution each time. The results obtained in this paper will be useful for exploring physical mechanisms of soliton excitations in nonlinear wave equations and applying the FNO in other nonlinear wave equations.

Keywords: deep learning, Fourier neural operator, solitonrogon state transition, nonlinear Schrödinger equation, Hirota equation

(Some figures may appear in colour only in the online journal)

1. Introduction

As a universal physical model, the focusing nonlinear Schrödinger (NLS) equation

$$i\partial_t\psi + \frac{1}{2}\partial_x^2\psi + |\psi|^2\psi = 0, \quad x, t \in \mathbb{R}^2, \quad (1)$$

where the amplitude envelope $\psi = \psi(x, t)$ is a complex function defined on the spatial location x and the evolution

time t , can be used to describe nonlinear wave phenomena of many conservative physical systems (e.g. fiber optics, plasma physics, marine science, Bose–Einstein condensates, finance, etc) [1–4]. It is also often referred to as the Gross–Pitaevskii (GP) equation in Bose–Einstein condensates [3]. Despite the simple expression form of the NLS equation, it contains a wealth of nonlinear wave structures [5, 6] (e.g. solitons, breathers, rogue waves, periodic solutions, etc). And the soliton-state transitions in the NLS equation can be realized by varying the parameters of explicitly exact solutions [7].

* Author to whom any correspondence should be addressed.

In recent years, due to the rise of the parity-time (\mathcal{PT}) symmetry [8], the NLS equation can also be extended to study the nonlinear waves in nonconservative systems [8–19]. And the nearly-integrable NLS equation with complex \mathcal{PT} -symmetric potentials can be written as [11–19]

$$i\partial_t\psi + \frac{1}{2}\partial_x^2\psi + |\psi|^2\psi + [V(x) + iW(x)]\psi = 0, \quad (2)$$

where the real-valued functions $V(x)$ and $W(x)$ denote the potential well and gain-loss distribution, respectively. The \mathcal{PT} -symmetry of the potential requires $V(x) = V(-x)$ and $W(x) = -W(-x)$. Due to the constraints of the dispersion term, the nonlinear term and the potential, clusters of solitons can also be found in equation (2). Thus the \mathcal{PT} -symmetry generalizes the concept of solitons, which were previously thought to exist only in conservative systems. Due to the complexity of the \mathcal{PT} -symmetric potentials, equation (2) also exists in the soliton-state transitions when the soliton parameters were varied in the \mathcal{PT} -symmetric NLS equation [16].

To study wave propagations of ultra-short optical pulses in Kerr nonlinear media, the focusing third-order NLS equation (alias Hirota equation) [20]

$$i\partial_t\psi + \frac{1}{2}\partial_x^2\psi + |\psi|^2\psi - i\beta(\partial_x^3\psi + 6|\psi|^2\partial_x\psi) = 0, \\ \beta \in \mathbb{R}, \quad (3)$$

was presented, which is also a fundamental physical model, and the third-order dispersion and higher-order nonlinear term are taken into account in the NLS equation (1). The Hirota equation (3) and its extensions can also be used to describe the strongly dispersive ion-acoustic wave in plasma physics [21] and the broader-banded waves in the deep ocean [22]. Particularly, equation (3) reduces to equation (1) at $\beta = 0$. The Hirota equation (3) is also completely integrable since it can be derived from the compatibility condition of the Lax pair [23, 24]. And it has been verified to admit the bright solitons [25–27], breathers [28, 29], and rogue waves [30–34] by means of the inverse scattering transform, Darboux transform, and numerical method. The state transitions from rogue waves to W-shaped solitons, breathers to the periodic waves, and breathers to anti-dark solitons have been investigated in [32, 33], which means that the solution parameters can be used to modulate the distinct wave structures of solitons.

Recently, deep learning (DL) [35] has attracted extensive attention due to the powerful fitting capability of neural networks. Meanwhile, the improvement of computer hardware (GPU) and the continuous updating of DL libraries (e.g. Pytorch [36], Tensorflow [37], Julia [38]) have also greatly promoted the development of DL. It is the continuous development of hardware and software that enables neural networks to process large amounts of data simultaneously and extract features for prediction. Nowadays, DL has impacted various areas, such as image recognition [39, 40], target detection [41], image generation [42], machine translation [43], speech recognition [44], natural language processing [45], and etc [46, 47]. The universal approximation theorem [48, 49] for neural networks has also created a boom in scientific machine learning (SciML) [50]. Due to its powerful

approximation ability, it is natural to apply it to the numerical approximation of ordinary differential equations and partial differential equations (PDEs).

The use of DL-based methods to solve differential equations (DEs) can be broadly divided into two genres. One is the neural finite element method, which replaces the finite element with a global neural network as the base of the approximation solution function of DE, such as the physics-informed neural networks (PINNs) method [51–53], the deep Ritz method [54]; the other is the DL architecture that learns the mapping (i.e. solution operator) between infinite-dimensional function spaces, such as the DeepOnet [55], Fourier neural operator (FNO) [56, 57], and other neural operator methods [58, 59]. The approximation errors of the two infinite-dimensional neural operators are proved theoretically in the latter papers [60, 61]. The most important feature of PINNs is the incorporation of the PDEs into the loss functions, which greatly increases the accuracies of the prediction solutions, but it requires retraining of the network for any given new instance of the function coefficients or initial conditions (see e.g. [62–69] and references therein). However, the prediction solutions via the FNO can be quickly given for any given new parameters once the networks were trained, while we do not need to know the underlying PDEs. In other words, we can unitedly learn the PDE families via the FNO. Thus, FNO has been extended to various fields, such as weather forecasting [70], the multi-phase flow [71], and heterogeneous material modeling [72]. If the structure of the temporal domain or spatial domain looks particularly complicated (high frequency) features to the frequency domain can be approximated with a small number of basis functions, then FNO can do convolution in the frequency domain with a small amount of data to learn out the operators between infinite dimensional spaces. These operators maintain the characteristics of Fourier approximation of high frequencies and neural network approximation of low frequencies [73, 74], so the FNO performs well in the approximation problems.

As mentioned above, the parameters of exact solutions can modulate the wave structures in some nonlinear integrable and nearly-integrable systems. In other words, we can obtain abundant nonlinear wave structures by changing the solution parameters. A natural and interesting question is how to use a deep neural network to uniformly learn these soliton-state transitions. To the aim, we combine the FNO with these soliton structures, and by training the FNO network we can obtain the different types of soliton solutions for a given initial condition, although it has a strong influence on the types of solitons. Moreover, we find that the soliton-state transitions of nonlinear wave equations can be well learned via the FNO network. To the best of our knowledge, the FNO was used to study the real-valued PDEs, but it was not extended to learn soliton-state transitions of the complex-valued nonlinear wave equations.

The rest of this paper is arranged as follows. In section 2, we specifically give the neural network framework of the FNO method, and the choice of some hyper-parameters. In section 3, we learn the different types of solitons in the NLS



Figure 1. Framework of Fourier neural operator (FNO), where Linear operator + Nonlinear integral operator + Activation function \rightarrow Highly learned nonlinear operator, and $v_{\tau+1}(x) = \sigma(Wv_{\tau}(x) + (\mathcal{K}(a; \Theta)v_{\tau})(x))$, $\tau = 0, 1, \dots, T - 1$, $x \in D$.

equation (1) via the data-driven FNO deep learning. In section 4, we learn the different types of solitons in the Hirota equation (3) via the data-driven FNO deep learning. The soliton state transitions of the NLS equation with \mathcal{PT} -symmetric potential given by equation (2) are discussed in section 5. Finally, we give some conclusions and discussions in section 6.

2. Deep learning-based FNO methodology

The FNO method [56, 57] aims to learn a nonlinear mapping $\mathcal{H}: \mathcal{A} \rightarrow \mathcal{U}$ between two infinite-dimensional function space \mathcal{A} and \mathcal{U} from the finite input-output data observations located in the bounded and open spatial-temporal domain $D = \{(x_i, t_i)\}_{i=1}^n \subset \mathcal{R}^2$, where \mathcal{A} and \mathcal{U} represent the parameter/initial condition spaces and (time-varying) solutions of a PDE, respectively. To approximate the operator, the FNO parameterizes \mathcal{H} to a parametric mapping \mathcal{H}_{Θ} via the neural network, which means

$$\mathcal{H}_{\Theta}: \mathcal{A} \times \Theta \rightarrow \mathcal{U} \tag{4}$$

with $\Theta: \mathcal{R} \rightarrow \mathcal{R}$ denoting the trainable finite-dimensional parameter space in the network. The input function $a(x) \in \mathcal{A}$ is first transformed into a higher-dimensional representation $v_0 = P(a(x))$ through a fully-connected shallow neural network $P(\cdot)$, then the frequency domain is learned by the following Fourier layers

$$v_{\tau+1}(x) = \sigma(Wv_{\tau}(x) + (\mathcal{K}(a; \Theta)v_{\tau})(x)), \tag{5}$$

$$\tau = 0, 1, \dots, T - 1,$$

where σ denotes a nonlinear activation function, W is the linear transform (i.e. weight) and $\mathcal{K}(a; \Theta)$ represents the kernel integral operator (i.e. kernel function) parameterized by the Θ , and can be defined as follow

$$(\mathcal{K}(a; \Theta)v_{\tau})(x) := \int_D \kappa(x, y, a(x), a(y); \Theta)v_{\tau}(y)dy, \tag{6}$$

where a kernel function κ denotes a neural network parameterized by Θ , and can be learned via data. Thus by supposing $\kappa(x, y; \Theta) = \kappa(x - y; \Theta)$ and convolution theorem, \mathcal{K} can be realized as

$$(\mathcal{K}(\Theta)v_k)(x) = \mathcal{F}^{-1}(R_{\Theta} \cdot (\mathcal{F}v_k))(x), \tag{7}$$

where R_{Θ} denotes the Fourier transform of a periodic function κ , and can be learned 1×1 convolution kernel (equivalent to the fully-connected layer between channels), \mathcal{F} and \mathcal{F}^{-1}

denote the Fourier transform and its inverse, respectively:

$$\mathcal{F}[f(x, t)](k) = \int_D f(x, t)e^{-2\pi ikx}dx,$$

$$\mathcal{F}^{-1}[\mathcal{F}[f]](k) = \int_D \mathcal{F}[f](k)e^{2\pi ikx}dk.$$

And in the realization, $\mathcal{F}, \mathcal{F}^{-1}$ can be obtained by the fast Fourier transform (FFT) of finite modes. It is the presence of the Fourier layer that enables the FNO to more quickly and accurately learn the high-frequency information, which in physical space corresponds to complex solutions. Finally, with a two-layer fully-connected neural network, the output can be obtained: $u(x, t)_{k=t_1}^T = Q(v_T(\cdot))$. That is, the output of the FNO can be written as the form

$$\mathcal{H}_{\Theta}(a(\cdot)) = Q \circ F_T \circ \dots \circ F_1 \circ P(a). \tag{8}$$

It is noted that we learn the mapping from the system parameters to the (time-varying) solution of the PDE in this paper. However, the constraint of FNO is that the inputs and outputs need to be defined on the same set of grids ($\{(x_i, t_i)\}_{i=1}^n$), except for the channels, thus we embody the changed parameters in the initial conditions of the PDE, i.e. $a(x) = u(x, t_0)$.

Training Data.—Since FNO is data-driven supervised learning, different from PINNs [51], it is essential to obtain the parameters of the whole network by high-quality label data. One can access the data from the exact solution (if available) or the numerical solutions via the high-accuracy numerical methods (e.g. spectral method, finite element method, etc). Also, for other problems in real life, the data can be derived from experiments and observations. In the following experiments, the parameters are generated from the uniform distribution on (a, b) , thus we can get the corresponding initial conditions. Though the parameters only vary from (a, b) , the impact on the types of solitons is huge which can be found in the discussions. If not specified, the total number of samples is 1500, of which 1000 are used for training and 200 for the test set.

The neural network framework.—Figure 1 displays the framework of the FNO. The fully-connected layer P increases the channels of the input function to width 32, then applies four Fourier integral operator layers with R being the 1×1 kernel. Finally, a two-layer fully-connected neural network Q changes the channel into 128 and the desired channel, which depends on the problem at hand. As for the finite term truncation of the FFT, we choose it as 12 in all problems. To overcome the internal covariate shift, the batch normalization [75] is applied to each Fourier layer. And the GeLu function $\sigma(x) = x\Phi(x)$ is chosen as the nonlinear activation function with $\Phi(x)$ denoting the distribution function of the standard normal distribution (notice that one can also choose other nonlinear activation functions). The Adam optimizer with the initial learning rate $\eta = 0.001$ halved every

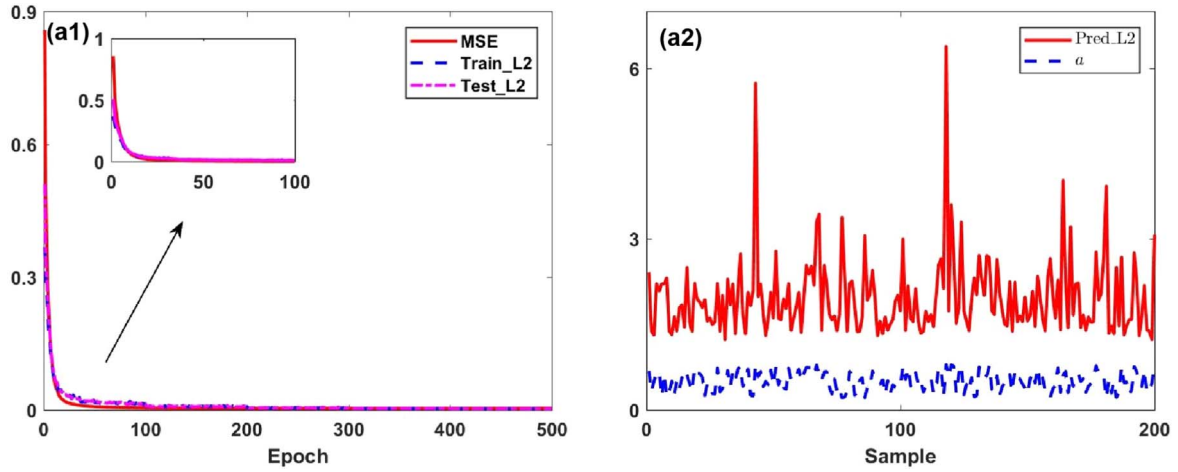


Figure 2. The training and testing progress for the problem (10). (a1) The mean squared error, the relative L_2 error on training (loss) and test sets versus epoch. (a2) The predicted L_2 error (10^{-3}) and sampled parameter a .

100 epochs in the total 500 epochs is used to minimize the loss function defined as

$$L(\psi, \hat{\psi}) = \frac{1}{N_{\text{train}}} \sum_{\ell=1}^{N_{\text{train}}} \frac{\|\psi_{\ell}(x, t) - \hat{\psi}_{\ell}(x, t)\|_2}{\|\psi_{\ell}(x, t)\|_2}, \quad (9)$$

with $\psi_{\ell}(x, t)$, $\hat{\psi}_{\ell}(x, t)$ representing the ground truth solution and the prediction solution of the ℓ th sample in train sets, respectively. And the batch size is chosen as 20 for the mini-batch stochastic gradient descent method. To observe the training process better, the mean squared error (MSE) and relative L_2 error on the test set are also recorded. It is worth noting that we treat the time t as a part of the spatial dimensions as well and apply FFT to it in the Fourier layer, so we have to add padding if necessary to preserve the periodic boundary conditions for the FFT. However, this does not mean that the FNO can only handle periodic problems due to the presence of Wv_{τ} in equation (5). It is worth noting that since the FNO is used to study complex nonlinear wave equations in this paper, we must add another channel in the implementation.

Remark. This is also the first application of the FNO method in complex-valued nonlinear wave equations such as the NLS equation (1), the \mathcal{PT} NLS equation (2) and the Hirota equation (3). Of course, the method can also be extended to other complex-valued nonlinear wave equations.

3. The deep learning soliton-state transitions of the NLS equation

In this section, we mainly use the data-driven FNO method to learn the soliton-state transitions of the NLS equation with the initial-value condition:

$$\begin{cases} i\partial_t \psi + \frac{1}{2} \partial_x^2 \psi + |\psi|^2 \psi = 0, & x \in (-6, 6), t \in (-6, 6), \\ \psi(x, -6) = \psi_0(x), & x \in [-6, 6]. \end{cases} \quad (10)$$

The NLS equation (10) admits the soliton solution with one parameter [5, 6]

$$\psi(x, t) = e^{it} \left[1 - \frac{2(1-2a)\cosh(bt) + ib \sinh(bt)}{\cosh(bt) - \sqrt{2a} \cos(\omega x)} \right], \quad (11)$$

where $b = [8a(1-2a)]^{1/2}$ and $\omega = 2(1-2a)^{1/2}$ with $a \in \mathbb{R}$. The only parameter a determines the different wave structures of this solution. As $0 < a < 0.5$, equation (11) denotes the Akhmediev breather (AB), the spatial-periodic, and temporal-local soliton, while it describes the Kuznetsov Ma (KM) breather when $a > 0.5$. And the rouge wave can be obtained from equation (11) by the limit of $a \rightarrow 0.5$:

$$\psi(x, t) = e^{it} \left[1 - \frac{4(1+2it)}{4(x^2+t^2)+1} \right]. \quad (12)$$

One can also observe the rouge wave-like solution in equation (11) with a being located nearby 0.5.

The spatial domain $[-6, 6]$ is split into 256 modes, while the temporal domain $[-6, 6]$ is split into 25 modes. Based on the problem, the 1500 samples are generated by selecting the uniform distribution of the parameter a on $(0, 1)$, then 1500 different initial conditions can be deduced. Here, we aim to learn the nonlinear operator mapping the initial condition $\psi(x, t_0)$ to the solution of some later time $T > t_0$, $\mathcal{H}: L^2([-6, 6], t = t_0; \mathbb{C}) \rightarrow L^2(x \in [-6, 6], t \in (t_0, T]; \mathbb{C})$ defined by

$$\mathcal{H}: \psi(x, t)|_{[-6,6],t=t_0} \rightarrow \psi(x, t)|_{[-6,6] \times (t_0,T]}. \quad (13)$$

In the implementation, the complex function $\psi(x, t)$ should be written into the real and imaginary parts, which is the first application of the FNO in complex nonlinear wave equations.

After the 500 training epochs, the parameters Θ in the neural networks can be obtained. And the total parameters in the neural networks is 2368 162. Finally, the MSE and the relative L_2 error on the train (loss) and test sets arrive $3.16\text{e-}3$, $1.93\text{e-}3$, $1.97\text{e-}3$, respectively. And the dynamic change versus epoch is depicted in figure 2(a1). As it can be seen that the loss function decreases rapidly in the first 50 epochs and then decreases at a

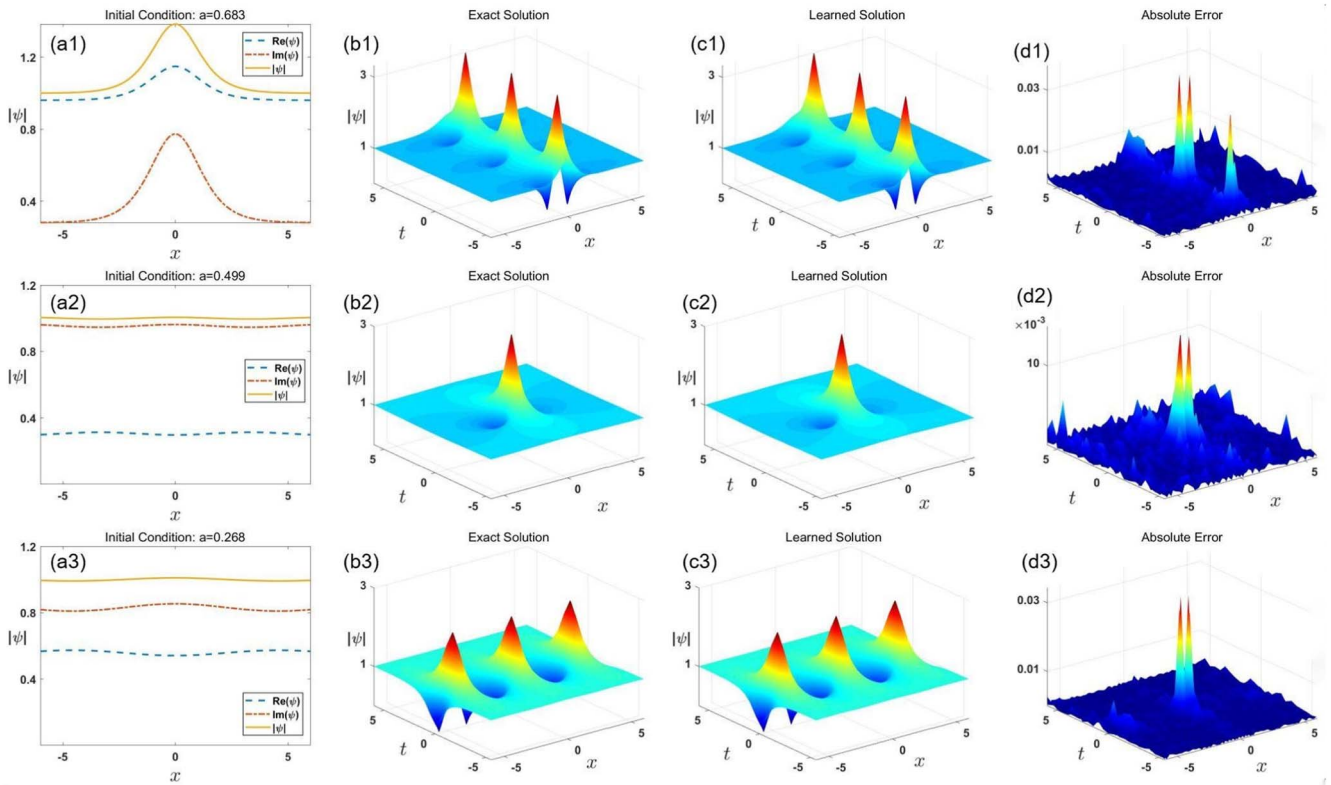


Figure 3. The data-driven soliton-state transitions of the NLS equation (10) via the FNO method. (a1), (b1), (c1), (d1) The Kuznetsov–Ma (KM) breather; (a2), (b2), (c2), (d2) Rouge waves; (a3), (b3), (c3), (d3) the Akhmediev breather (AB). The first column indicates the parameters and the corresponding initial conditions, the second column indicates the exact solutions, the third column is the learned solitons via the FNO, and the fourth column is the absolute errors.

very slow rate. It can be observed that the FNO performs well in the test set, though the initial condition is a new instance for the neural networks from figure 2(b). It is noted that the predicted L_2 error is 10^{-3} units displayed in figure 2(a2).

The different types of data-driven solutions via FNO belonging to the different parameters are exhibited in what follows (see figure 3):

Case 1. As for the Kuznetsov–Ma breather, we choose $a = 0.683$ in the test set. The corresponding initial condition is presented in figure 3(a1). It can be concluded that the learned soliton via FNO matches well with the exact solution from figures 3(b1), (c1), (d1). It is worth mentioning that we find that the approximation error of PINNs is similar in magnitude to that of FNOs, however, the FNO can learn the entire family of solutions under a given distribution, while the PINNs can only learn some specific solution each time. This illustrates the superiority of the neural operator method.

Case 2. For the rouge waves, figure 3(a2) displays the parameter and initial condition, which differs from the KM breather. Though the coefficients/initial conditions have a strong influence on the type of solution, FNO still demonstrates its strong approximation capability. It can be seen that

the learned FNO rouge wave is very close to that of the exact solution from figures 3(b2), (c2), (d2).

Case 3. We also choose $a = 0.268 < 0.5$ and the initial condition in figure 3(a3) to depict the predicted structure of spatial-periodic, temporal-local soliton. In figures 3(b3), (c3), (d3), the simulation results demonstrate the high accuracy of the Akhmediev breather.

From the above results, we can conclude that the FNO performs well in the soliton state transitions of the NLS equation (10) by varying the parameters/initial conditions. Also, the FNO achieves state-of-the-art (SOTA) in terms of prediction accuracy of the solutions although the dynamical behaviors of the solitons are more complicated.

4. The deep learning soliton-state transitions of the Hirota equation

In this section, we mainly use the data-driven FNO method to deep learn the soliton-state transitions of the Hirota equation (3) with the initial-value condition:

$$\begin{cases} i\partial_t \psi + \frac{1}{2}\partial_x^2 \psi + |\psi|^2 \psi - i\beta(\partial_x^3 \psi + 6|\psi|^2 \partial_x \psi) = 0, & x \in (-8, 8), t \in (-6, 6), \\ \psi(x, -6) = \psi_0(x), & x \in [-8, 8]. \end{cases} \quad (14)$$

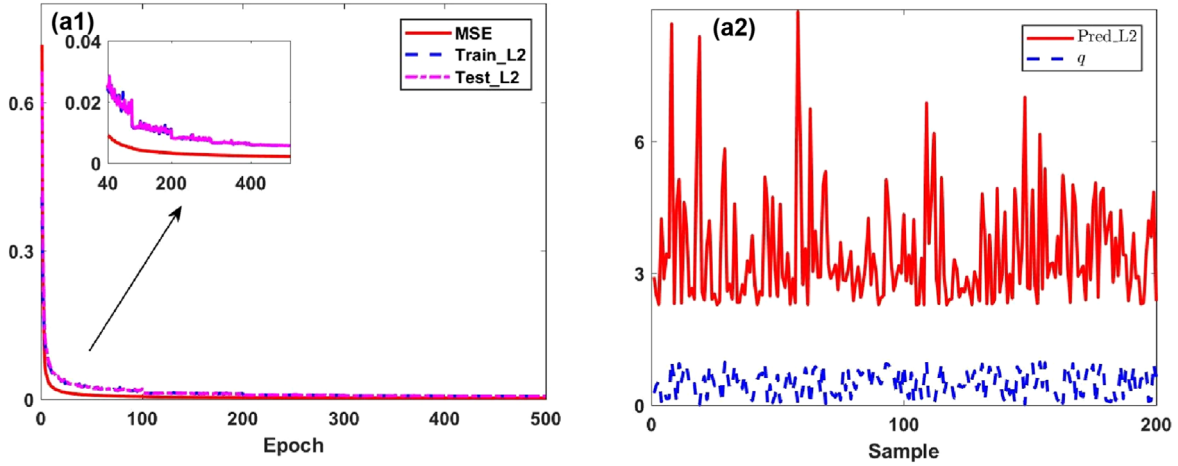


Figure 4. The training and testing progress on equation (14). (a1) The mean squared error, the relative L_2 error on training (loss) and test sets versus epoch. (a2) The predicted L_2 error (10^{-3}) and sampled parameter a .

4.1. Learning an operator from rouge waves to W-shaped solitons

The solutions of the Hirota equation (14) can be deduced as

$$\psi(x, t) = \kappa_0 \left[\frac{4 + 8ia^2(1 - q/q_s)\tau}{1 + 4a^2(\xi - v\tau)^2 + 4a^4(1 - q/q_s)^2\tau^2} - 1 \right], \quad (15)$$

with κ_0 being a non-zero background solution

$$\kappa_0 = ae^{i\theta}, \quad \theta = qx + [a^2 - q^2/2 + \beta(6qa^2 - q^3)]t \quad (16)$$

by means of Darboux transformation method, where $v = q + (2a^2 - q^2)/(2q_s)$, $\xi = x - x_c$, $\tau = t - t_c$, and (x_c, t_c) determines the center of the solution. And a and q denote the amplitude and wavenumber of the background solution, respectively. The state transitions induced by the higher-order effects and background frequency have been analyzed via the modulation instability [32]. For the case $q \neq q_s$, equation (15) admits two types of rouge waves with different dynamic behaviours ($q = 0, q_s/2$). However, in the case $q = q_s$, equation (15) reads the W-shaped waves with one peak and two valleys

$$\psi_s(x, t) = ae^{i\theta_s} \left[\frac{4}{1 + 4a^2(\xi - v_s\tau)^2} - 1 \right], \quad (17)$$

with $\theta_s = \theta|_{q=q_s}$, $v_s = v|_{q=q_s}$. The W-shaped soliton (17) can also be obtained by choosing $q \rightarrow 1$ in equation (16).

We devote ourselves to researching the state-transition from the rouge wave to a W-shaped soliton in the Hirota equation (14) by modulating the parameter. Thus, the FNO can be used to learn the state transitions in equation (15). Without loss of generality, we choose $q_s = 1$, $t_c = 3$, $x_c = 0$, $a = 1$, and $\beta = 0.1$ in the following discussions. The spatial domain $[-8, 8]$ is split into 256 modes, while the temporal domain $[-6, 6]$ is discrete into 49 modes for the input-output observations. It is worth mentioning that we have added a fine time split to capture the structure of the solution more precisely here. Considering the effect of q/q_s on the solution pattern, we select 1500 samples uniformly distributed on $(0, 1)$ and

substitute them into equation (15) to obtain the corresponding initial conditions. Similarly to the previous case, although the fluctuation of q varies tinily, it has a huge influence on the initial condition and the localized waves at later moments. Our goal here is to use the FNO to learn the mapping from the initial condition $\psi(x, t_0)$ to the solution of some later time $T > t_0$, $\mathcal{H}: L^2([-8, 8]; t = t_0; \mathbb{C}) \rightarrow L^2([-8, 8]; t \in (t_0, T]; \mathbb{C})$ defined by

$$\mathcal{H}: \psi(x, t)|_{[-8, 8], t=t_0} \rightarrow \psi(x, t)|_{[-8, 8] \times (t_0, T]}. \quad (18)$$

The real and imaginary parts need to be separated in order to carry out the output of the input of the network.

The FNO can be trained by minimizing the relative L_2 error for 500 epochs via the mini-batch gradient descent using the Adam optimizer with a decay learning rate. It is worth mentioning that the time of each epoch increases to about 4.1s due to the increase of the time discretization. Once the network is trained, the network can give the corresponding solution $\psi(x, t) \in \mathcal{U}$ in a very short time for a new input instance $a \in \mathcal{A}$. During the training process, we record the mean squared error, the relative L_2 error on train, and test sets. The dynamic changes versus epoch is depicted in figure 4(a1). The values arrive 2.49e-3, 3.33e-3, 3.44e-3, respectively. The values of the loss functions decrease faster in the first 300 epochs, and then slower in the last 200 epochs. We can find that the FNO also performs very well on the test set from figure 4(a2). The relative L_2 errors of the learned solutions are mostly below 10^{-2} .

To better visualize the type of travelling waves, we also choose three parameters/initial conditions (new instances do not appear in the training set) to demonstrate the processes of soliton state-transitions (see figure 5):

Case 1. Firstly, we choose $q = 0.065$ to display the rouge waves, and the corresponding initial condition is exhibited in figure 5(a1). The dynamic behaviors of the learned solution, the exact solution, and the absolute error between them are demonstrated in figures 5(b1)–(d1). The learned solution captures the characteristics of the rouge wave well, and the maximum absolute error is only around 0.05.

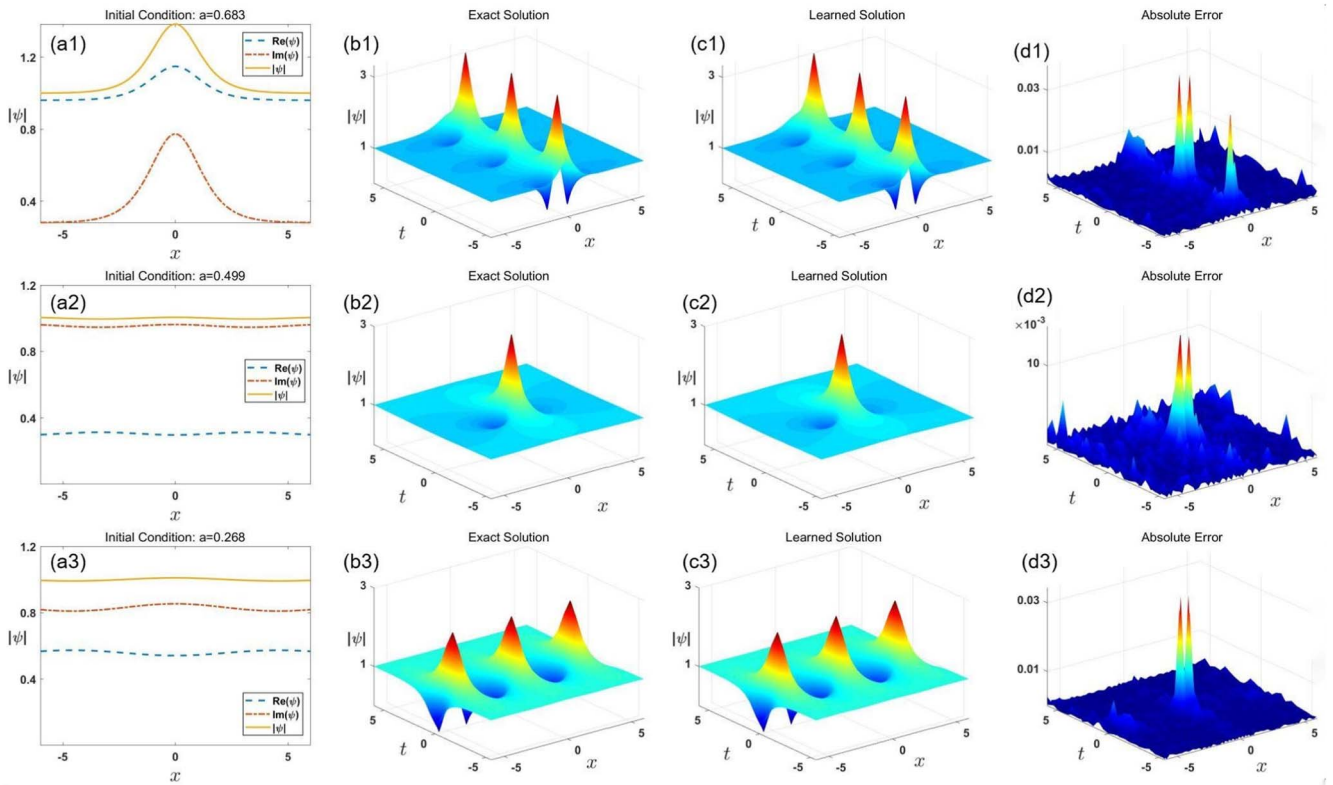


Figure 5. The data-driven soliton-state transitions of solutions (15) in the Hirota equation (14) via FNO. (a1)–(d1): the rouge wave; (a2)–(d2): the elongated rouge wave; (a3)–(d3): the W-shaped soliton with one peakon and two valleys. The first column indicates the parameter and the corresponding initial condition, the second column indicates the exact solution, the third column is the learned soliton via FNO, and the fourth column is the absolute error.

Case 2. We then set $q = 0.530$ to explore the features of the elongated rouge waves (see figure 5(a2)). The representative visualizations of the predicted solutions for the input samples are shown in figures 5(b2)–(d2). As it can be seen that the prediction achieves a good agreement with exact soliton over the entire spatial-temporal domain. It is worth noting that the initial condition, in this case, is similar to that in figure 5(a1), but the solution at a later time is different. However, FNO still captures the two different types of solutions well, which is a good indication of the superiority of the FNO algorithm for the approximations of complex problems.

Case 3. For the case of W-shaped travelling waves, $q = 0.973$ and the initial condition are chosen as the input of the deep neural network. We can observe that the input for the network differs from the case of the above two cases from figure 5(a3). It can be seen that an excellent agreement can be realized between the FNO’s learned solutions and the exact solitons in figures 5(b3)–(d3). Despite the strong influence of the parameters on the solution, the predicted solutions via FNO can be learned very well.

Based on the above results, the Fourier layer in the FNO plays a critical role in the network to capture the high frequency part. That is, the FNO is able to distinguish well between different frequencies although the inputs (initial values) are very similar. And this is what is missing in traditional neural networks [73, 74].

4.2. Learning an operator from Akhmediev breathers to periodic waves

Here we study the state transition from the Akhmediev breather to the periodic wave and the related wave phenomena. Equation (3) also possesses the Akhmediev breather in the form [33]

$$\psi(x, t) = \kappa_0 \left[\frac{2\eta^2 \cosh(\kappa t) + 2i\eta a_1 \sinh(\kappa t)}{\cosh(\kappa x) - a_1/a \cos[2\eta(x + v_1 t)]} - 1 \right], \quad (19)$$

where κ_0 is given by equation (16), $v_1 = \beta(2a^2 + 4a_1^2 - q^2) - 2q(q\beta + 1/2)$, $\kappa = 2\eta v_2$, $\eta = \sqrt{a^2 - a_1^2}$, $v_2 = a_1(1 - q/q_s)$, $q_s = -1/(6\beta)$. Similarly, the shape of equation (19) is changed with q/q_s as $q \neq q_s$ ($q = 0$, $q = q_s/2$). As for $q = q_s$, the Akhmediev breather (19) is converted to the periodic wave

$$\psi(x, t) = \kappa_0 \left[\frac{2\eta^2}{1 - a_1/a \cos[2\eta(x + vt)]} - 1 \right], \quad (20)$$

which can also be obtained by the limit $q \rightarrow q_s$ in equation (19).

To this end, the FNO is extended to learn the conversion from the Akhmediev breather to the periodic wave by varying q (initial condition) in equation (19). The parameters are chosen as $a = 1$, $a_1 = 0.7$, $\beta = 0.1$ in the numerical experiments. Since the structure of the wave is more complicated at

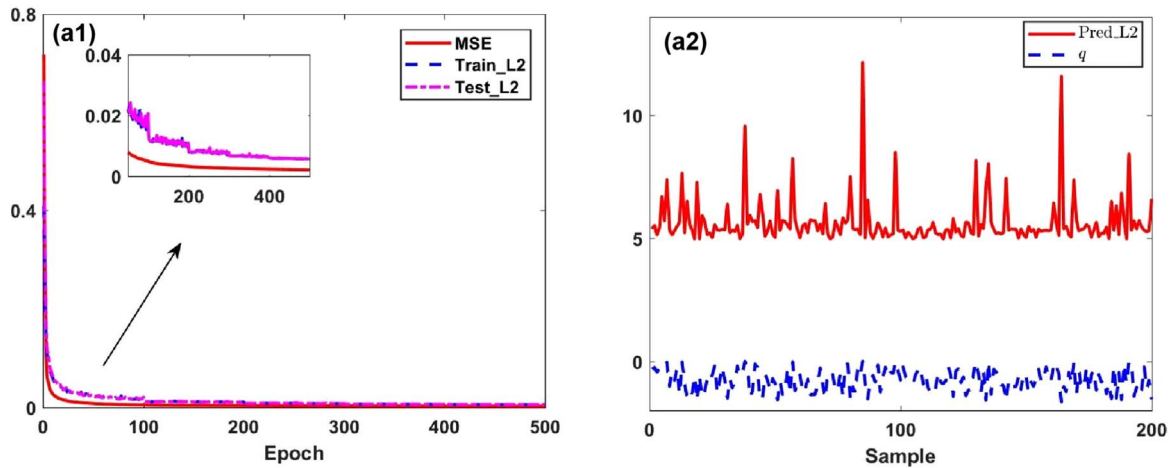


Figure 6. The training and testing progress of equation (19). (a1) The mean squared error, the relative L_2 error on training (loss) and test sets versus epoch. (a2) The predicted L_2 error (10^{-3}) and sampled parameter q .

this point, we increase the number of divisions in the spatial-temporal domain. We discrete $(x, t) \in [-10, 10] \times [-6, 6]$ into 512×121 grids for the finite collections of input-output pairs. the FNO aims to learn the mapping from the initial condition $\psi(x, t_0)$ to the solution of some later time $T > t_0$, $\mathcal{H}: L^2([-10, 10]; t = t_0; \mathbb{C}) \rightarrow L^2([-10, 10]; t \in (t_0, T]; \mathbb{C})$ defined by

$$\mathcal{H}: \psi(x, t)|_{[-10,10],t=t_0} \rightarrow \psi(x, t)|_{[-10,10] \times (t_0,T]}. \quad (21)$$

Since the neural network only supports the real-valued input and output, it is necessary to separate the initial conditions into real and imaginary parts for the input, and finally combine the real and imaginary parts of the output to get the prediction solutions.

By training the network for 500 epochs with the Adam optimizer, the parameters Θ can be obtained by the gradient descent method and back propagation mechanisms. Due to the increase in the number of configured points in the spatial-temporal domain, the time to run each epoch also increases to about 17s. However, once the network is trained, it can make predictions in less than 1s for each instance on the test set, which is much faster than the traditional algorithms. To better visualize the training process, we record the MSE, relative L_2 -error on train, and test sets. These values eventually arrive $2.23e-3$, $5.79e-3$, $5.64e-3$, respectively. It can be observed that the value of the loss function decreases faster in the first 200 epochs and slower in the last 300 epochs from figure 6(a1). The performance of FNO on the test set is also recorded in figure 6(a2). From the relative L_2 error, we can see that the FNO still behaves well on 200 unseen samples.

In order to display the change of the soliton state, we choose a few specific parameters (initial conditions) located in the test set (see figure 7):

Case 1. In the case of the Akhmediev breather, $q = -10^{-3}$ is chosen to exhibit the initial condition in figure 7(a1). Figures 7(b1)–(d1) also demonstrate the FNO’s learned soliton, ground truth, and the absolute error between the two. Despite the complicated structure of the Akhmediev breather, the FNO can still capture these features very well.

Meanwhile, the absolute error of 0.1 is achieved only in a few positions.

Case 2. Then $q = -0.830$ is utilized to explore the decreased t -direction localization Akhmediev breather, the corresponding initial condition is depicted in figure 7(a2). As it can be seen that although the value of q does not fluctuate much, the effect on the initial value is more significant. The excellent agreement between the predicted soliton and the exact solution is displayed in figures 7(b2-d2) while the maximum error is only around 0.05. Surprisingly, the FNO is able to learn complicated structures for different initial values while ensuring the accuracy of the predicted solutions.

Case 3. Finally, we set $q = -1.659$ to learn the corresponding periodic wave (see figure 7(a3)). As shown in figures 7(b3)–(d3), the predicted solutions are in qualitative agreement with the corresponding ground truths.

On the basis of the above results, we can draw the conclusion that for the initial conditional input $a \in \mathcal{A}$ with a given distribution, the FNO is able to extract the key features with only 1000 samples, although there is a large gap in the ground truth corresponding to these 1000 samples. Meanwhile, the network is able to give the output $\psi(x, t) \in \mathcal{U}$ for new samples quickly which greatly improves the generalization ability of FNO. Surprisingly, compared with the traditional neural network prediction model, the solutions learned by FNO also achieve SOTA in terms of accuracy.

5. The NLS equation with the complex \mathcal{PT} -symmetric potential

In this section, we mainly consider the data-driven FNO deep learning of the soliton-state transitions of the NLS equation with \mathcal{PT} -symmetric potential given by equation (2) with the initial-value condition:

$$\begin{aligned} i\partial_t\psi + \frac{1}{2}\partial_x^2\psi + |\psi|^2\psi + [V(x) + iW(x)]\psi &= 0, \\ x \in (-10, 10), t \in (0, 5), \\ \psi(x, 0) = \psi_0(x), \quad x \in [-10, 10], \end{aligned} \quad (22)$$

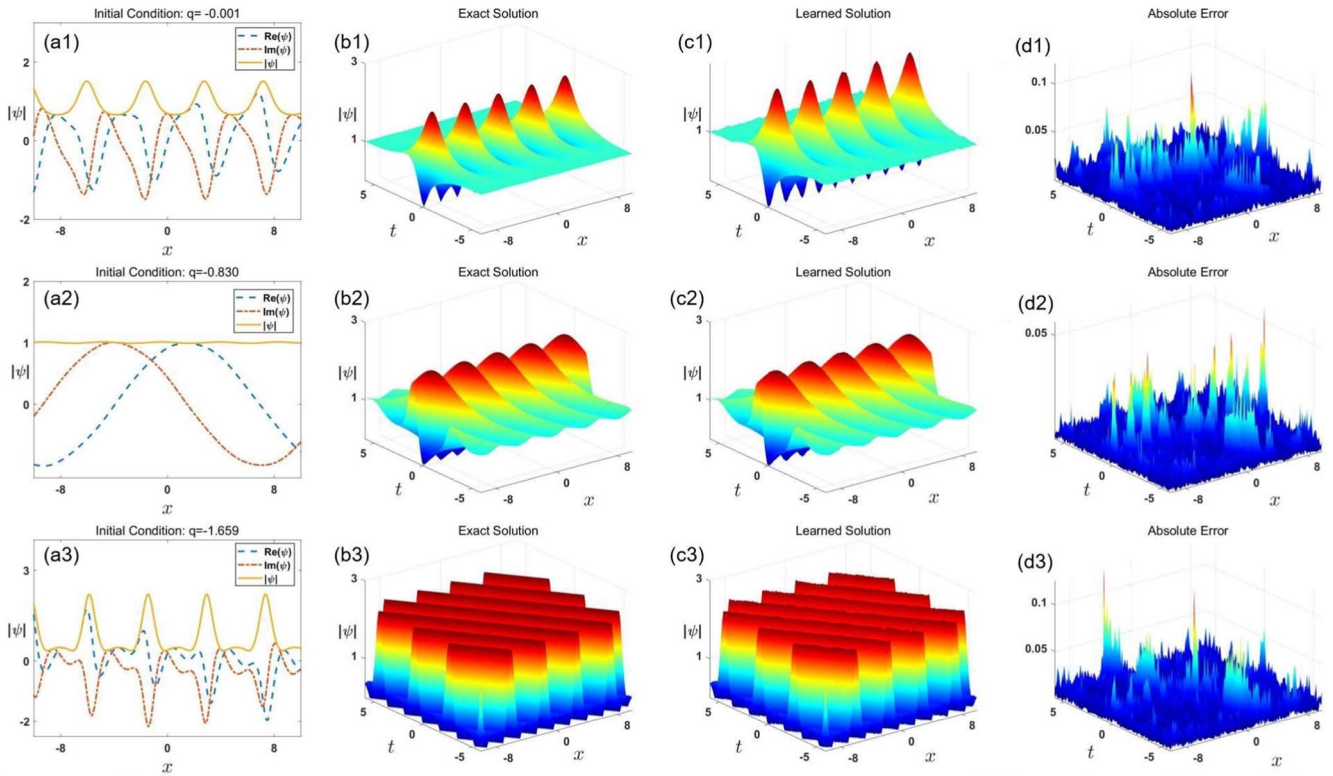


Figure 7. The data-driven soliton-state transitions of solutions (19) in the Hirota equation (14) via FNO. (a1–d1): the Akhmediev breather; (a2–d2): the decreased t -direction localization Akhmediev breather; (a3–d3): the periodic waves. The first column indicates the parameter and corresponding initial condition, the second column indicates the exact solution, the third column is the learned soliton via FNO, and the fourth column is the absolute error.

where the complex potential is chosen as the generalized \mathcal{PT} -symmetric Scarf-II ($\delta(x)$ -Scarf-II) potential [16]

$$\begin{aligned} V(x) &= 2\alpha(\delta(x) - \tanh|x|) - 2 \operatorname{sech}^2 x, \\ W(x) &= -W_0 \partial_x(\operatorname{sech} x e^{\alpha|x|}), \quad \alpha < 1, \end{aligned} \quad (23)$$

which becomes the \mathcal{PT} -symmetric Scarf-II potential at $\alpha = 0$ [9, 11].

In particular, equation (22) with the \mathcal{PT} potential (23) can be shown to possess the solitons

$$\begin{aligned} \psi(x, t) &= \phi(x) e^{i\mu t}, \quad \phi(x) = \frac{W_0}{3} \operatorname{sech}(x) e^{\alpha|x|} \\ &\exp\left(-\frac{iW_0}{3} \int_0^x \operatorname{sech}(s) e^{\alpha|s|} ds\right), \end{aligned} \quad (24)$$

and $\mu = 1 + \alpha^2$. It follows from solution (24) that equation (22) possesses:

- The peakon solution when $\alpha < 0$.
- The smooth bright soliton for $\alpha = 0$.
- The double-hump peakon solution with one sharp valley if $0 < \alpha < 1$.

Similarly, the soliton-state transitions can be modulated by the system parameter α , thus FNO can be applied to find the abundant structures in the \mathcal{PT} -symmetric NLS equation. We utilize 256 modes and 6 modes to discrete the spatial-temporal domain $(x, t) \in [-10, 10] \times [0, 5]$ for the finite observations of $a_j(x) \in \mathcal{A}$ and $u_j(x) \in \mathcal{U}$, where j indicates the input-output pairs. Due to the effect of the

parameter α on the type of solitons, we choose 1500 α uniformly distributed on $(-1, 1)$ and generate 1500 different types of initial conditions through it as the FNO’s input. We are interested in learning the operator mapping the initial condition $\psi(x, 0)$ to the soliton up to some later time $T > 0$, $\mathcal{H}: L^2([-10, 10]; t = 0; \mathbb{C}) \rightarrow L^2([-10, 10]; t \in (0, T]; \mathbb{C})$ defined by

$$\mathcal{H}: \psi(x, t)|_{[-10, 10], t=0} \rightarrow \psi(x, t)|_{[-10, 10] \times (0, T]}. \quad (25)$$

Similar to the previous implementation, we add an extra channel to separate the real and imaginary parts of the complex function $\psi(x, t)$.

The parameters Θ can be obtained after training the FNO by minimizing the loss function defined by equation (9) for 500 epochs with the mini-batch gradient descent method. The results for the mean squared error, relative L_2 error on train, and test sets in the process are presented in figure 8(a1). Throughout the training process, the Loss function decreases in a fast and then slow trend. And the values finally arrive $9.2e-4, 4.78e-3, 4.81e-3$, respectively. The prediction L_2 error is measured in new unseen instances that are not used during the training process which is exhibited in figure 8(a2). As can be seen, the errors on the 200 test samples are most less than 10^{-2} .

Furthermore, we investigate the predicted solution, the ground truth as well as the absolute error under the different

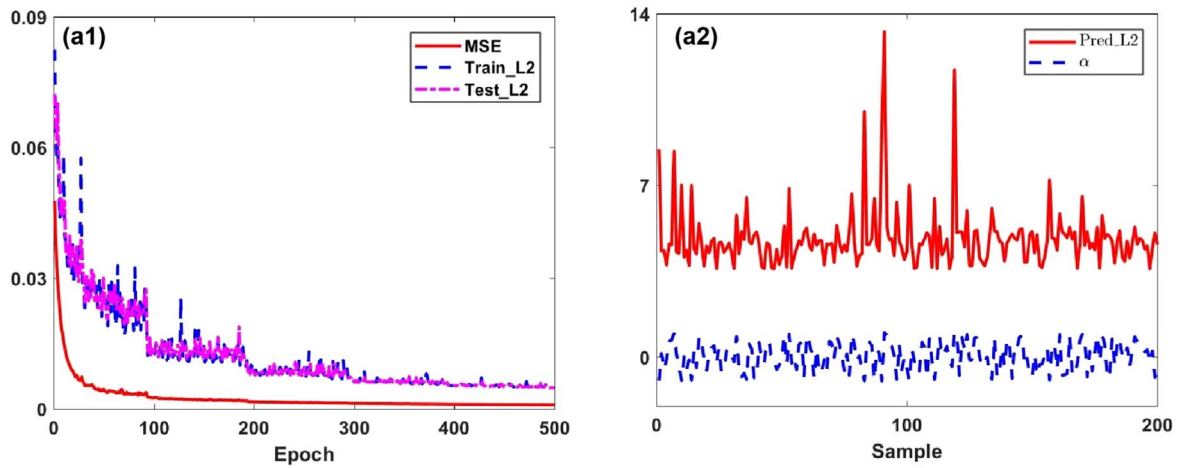


Figure 8. The training and testing progress of equation (24). (a1) The mean squared error, the relative L_2 error on training (loss) and test sets versus epoch. (a2) The predicted L_2 error (10^{-3}) and sampled parameter α .

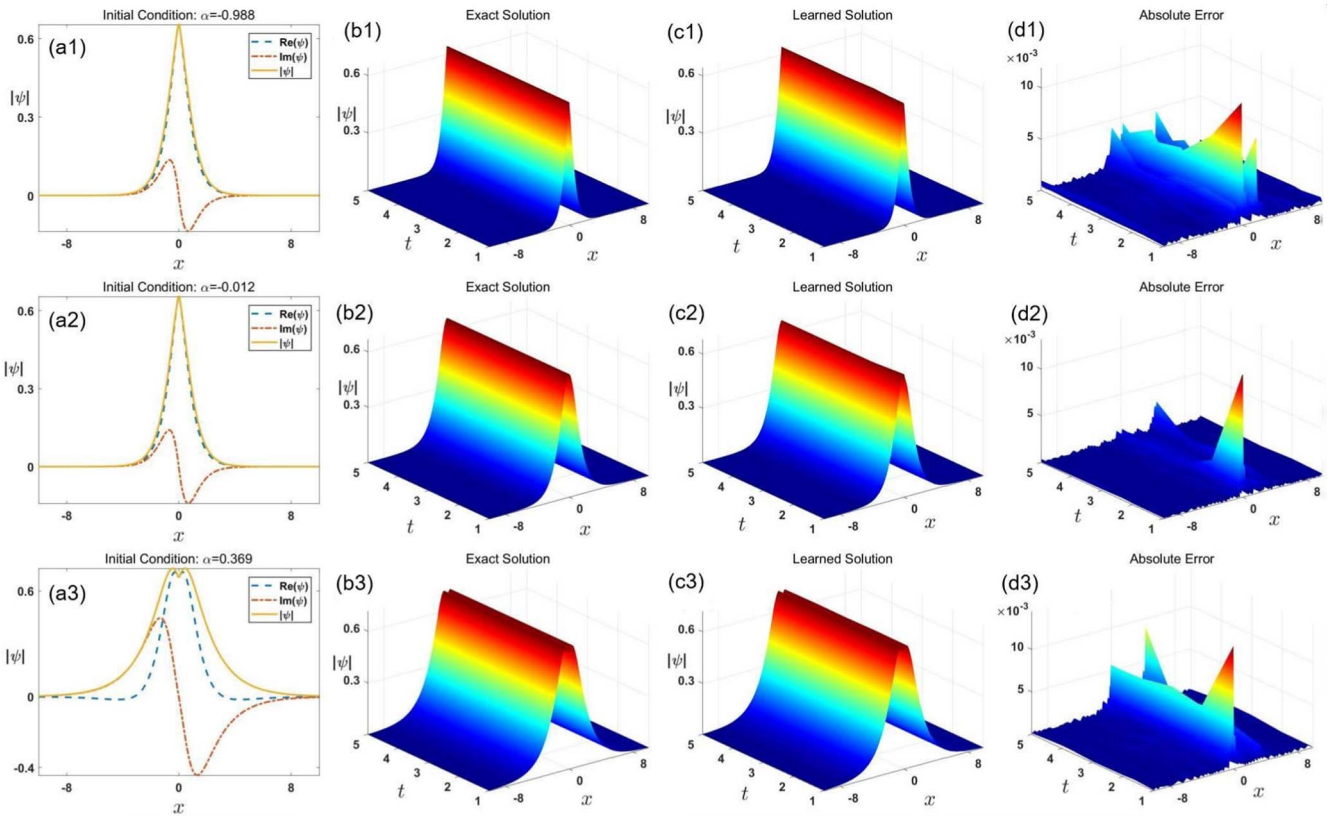


Figure 9. The data-driven soliton-state transitions of equation (24) with $W_0 = 2$ and varying α in the \mathcal{PT} -symmetric NLS equation (22) via FNO. (a1)–(d1): the peakon soliton; (a2)–(d2): the smooth bright soliton; (a3)–(d3): the double-hump soliton with one sharp valley. The first column indicates the parameter and the corresponding initial condition, the second column indicates the exact solution, the third column is the learned solution via FNO, and the fourth column is the absolute error.

initial conditions for visualization (see figure 9, where $W_0 = 2$):

Case 1. We first consider the peakon soliton given by equation (24) with $\alpha < 0$, which is not differentiable at $x = 0$. $\alpha = -0.988$ and the corresponding initial condition for the input sample is depicted in figure 9(a1). And figures 9(b1–d1) display the comparison between the exact soliton and the learned solution for the input. We can observe that the FNO’s

predicted solution achieves an excellent agreement with the ground truth.

Case 2. We also choose $\alpha = -0.012$ for the presentation of smooth bright soliton in figure 9(a2). It is evident that the learned solution matches well with the ground truth from the visualizations exhibited in figures 9(b2)–(d2).

Case 3. The double-hump soliton with one sharp valley is also investigated by setting $\alpha = 0.369$ which is unseen in

the training sets (see figure 9(a3)). Thanks to the transformation of high frequency information by the Fourier layer, it can be seen that the predicted solution of FNO captures the small structure of the complex solution very well (one sharp valley) depicted in figures 9(b3)–(d3) while the high accuracies are also maintained.

On the basis of these results, we can conclude that the FNO can be regarded as a powerful data-driven deep learning method that the generalization ability of the network is greatly enhanced compared to the PINNs [52], which the input parameters/initial conditions are given and fixed during the training and testing process. Besides, the FNO can predict solitons with more complicated wave structures.

6. Conclusions and discussions

In summary, we have effectively studied the soliton state-transitions in some soliton equations by means of the FNO method. We consider various types of nonlinear wave excitations in the NLS equation, Hirota equation, and NLS equation with PT -symmetric potential. Despite the complicated structures of the solutions, the FNO is still able to capture the key features and information. And for the given new samples, it can give more accurate prediction solutions. The results obtained in this paper are useful to further analyze the soliton-state transfers and neural operator networks solving complex-valued nonlinear PDEs.

In the case that the studied equation has no analytical solution, one can find its numerical solutions. Since FNO is a supervised deep learning model, the quality of supervised data will affect the approximation effect of FNO to a certain extent. Therefore, we need to obtain supervised data through some high-precision numerical algorithm as the input label. However, due to the limitations of numerical algorithms, it is sometimes difficult to obtain supervised data, so it is particularly important how to design an unsupervised learning operator method. This will be considered and discussed in our future work.

Acknowledgments

The work of ZY was supported by the NSFC under Grant Nos. 11925108 and 11731014. The work of S-FT was supported by the NSFC under Grant No. 11975306.

References

- [1] Ablowitz M J and Clarkson P A 1991 *Solitons, Nonlinear Evolution Equations and Inverse Scattering* (Cambridge: Cambridge University Press)
- [2] Kharif C, Pelinovsky E and Slunyaev A 2009 *Rogue Waves in the Ocean* (Berlin: Springer)
- [3] Pitaevskii L and Stringari S 2016 *Bose–Einstein Condensation and Superfluidity* (Oxford: Oxford University Press)
- [4] Yan Z 2010 Financial rogue waves *Commun. Theor. Phys.* **54** 947
- [5] Akhmediev N and Kornev V I 1986 Modulation instability and periodic solutions of the nonlinear Schrödinger equation *Theor. Math. Phys.* **69** 1089
- [6] Peregrine D H 1983 Water waves, nonlinear Schrödinger equations and their solutions *J. Aust. Math. Soc. B* **25** 16
- [7] Toenger S, Godin T, Billet C, Dias F, Erkintalo M, Genty G and Dudley J M 2015 Emergent rogue wave structures and statistics in spontaneous modulation instability *Sci. Rep.* **5** 1
- [8] Bender C M and Boettcher S 1998 Real spectra in non-Hermitian Hamiltonians having PT symmetry *Phys. Rev. Lett.* **80** 5243
- [9] Ahmed Z 2001 Real and complex discrete eigenvalues in an exactly solvable one-dimensional complex PT -invariant potential *Phys. Lett. A* **282** 343
- [10] Bender C M 2007 Making sense of non-Hermitian Hamiltonians *Rep. Prog. Phys.* **70** 947
- [11] Musslimani Z, Makris K G, El-Ganainy R and Christodoulides D N 2008 Optical solitons in PT periodic potentials *Phys. Rev. Lett.* **100** 030402
- [12] Kevrekidis P G, Cuevas-Maraver J, Saxena A and Cooper F 2015 Interplay between parity-time symmetry, supersymmetry, and nonlinearity: an analytically tractable case example *Phys. Rev. E* **92** 042901
- [13] Yan Z, Wen Z and Hang C 2015 Spatial solitons and stability in self-focusing and defocusing Kerr nonlinear media with generalized parity-time-symmetric Scarf-II potentials *Phys. Rev. E* **92** 022913
- [14] Yan Z, Wen Z and Konotop V V 2015 Solitons in a nonlinear Schrödinger equation with PT -symmetric potentials and inhomogeneous nonlinearity: Stability and excitation of nonlinear modes *Phys. Rev. A* **92** 023821
- [15] Chen Y, Yan Z, Mihalache D and Malomed B A 2017 Families of stable solitons and excitations in the PT -symmetric nonlinear Schrödinger equations with position-dependent effective masses *Sci. Rep.* **7** 1
- [16] Zhong M, Chen Y, Yan Z and Tian S-F 2022 Formation, stability, and adiabatic excitation of peakons and double-hump solitons in parity-time-symmetric Dirac- $\delta(x)$ -Scarf-II optical potentials *Phys. Rev. E* **105** 014204
- [17] Moiseyev N 2011 *Non-Hermitian Quantum Mechanics* (New York: Cambridge University Press)
- [18] Konotop V V, Yang J and Zezyulin D A 2016 Nonlinear waves in PT -symmetric systems *Rev. Mod. Phys.* **88** 035002
- [19] Christodoulides D N and Yang J 2018 *Parity-time Symmetry and its Applications* (New York: Springer)
- [20] Hirota R 1973 Exact envelope soliton solutions of a nonlinear wave equation *J. Math. Phys.* **14** 805
- [21] Trulsen K and Dysthe K B 1996 A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water *Wave Motion* **24** 281
- [22] Craig W, Guyenne P and Sulem C 2012 Hamiltonian higher-order nonlinear Schrödinger equations for broader-banded waves on deep water *Eur. J. Mech. B* **32** 22
- [23] Ablowitz M J, Kaup D J, Newell A C and Segur H 1973 Nonlinear-evolution equations of physical significance *Phys. Rev. Lett.* **31** 125
- [24] Chen H H 1974 General derivation of Bäcklund transformations from inverse scattering problems *Phys. Rev. Lett.* **33** 925
- [25] Lakshmanan M and Ganesan S 1983 Equivalent forms of a generalized Hirota's equation with linear inhomogeneities *J. Phys. Soc. Jpn.* **52** 4031
- [26] Mihalache D, Torner L, Moldoveanu F, Panoiu N C and Truta N 1993 Inverse-scattering approach to femtosecond solitons in monomode optical fibers *Phys. Rev. E* **48** 4699
- [27] Porseizian K and Nakkeeran K 1996 Optical solitons in presence of Kerr dispersion and self-frequency shift *Phys. Rev. Lett.* **76** 3955

- [28] Li L, Li Z, Xu Z, Zhou G and Spatschek K H 2002 Gray optical dips in the subpicosecond regime *Phys. Rev. E* **66** 046616
- [29] Li S Q, Li L, Li Z H and Zhou G S 2004 Properties of soliton solutions on a cw background in optical fibers with higher-order effects *J. Opt. Soc. Am. B* **21** 2089
- [30] Ankiewicz A, Soto-Crespo J M and Akhmediev N 2010 Rogue waves and rational solutions of the Hirota equation *Phys. Rev. E* **81** 046602
- [31] Li L, Wu Z, Wang L and He J 2013 High-order rogue waves for the Hirota equation *Ann. Phys.* **334** 198
- [32] Liu C, Yang Z Y, Zhao L C and Yang W L 2015 State transition induced by higher-order effects and background frequency *Phys. Rev. E* **91** 022904
- [33] Liu C, Yang Z Y, Zhao L C, Duan L, Yang G and Yang W L 2016 Symmetric and asymmetric optical multipeak solitons on a continuous wave background in the femtosecond regime *Phys. Rev. E* **94** 042221
- [34] Wang L, Yan Z and Guo B 2020 Numerical analysis of the Hirota equation: modulational instability, breathers, rogue waves, and interactions *Chaos* **30** 013114
- [35] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436
- [36] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Proc. Adv. Neural Inf. Process. Syst.* **32** 8024
- [37] Abadi M *et al* 2016 TensorFlow: a system for large-scale machine learning *Proc. XII USENIX Symp. on Operating Systems Design and Implementation (OSDI)* p 265
- [38] Bezanson J, Edelman A, Karpinski S and Shah V B 2017 Julia: a fresh approach to numerical computing *SIAM Rev.* **59** 65
- [39] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks *Proc. Adv. Neural Inf. Process. Syst.* **25** 1097
- [40] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* p 770
- [41] Girshick R 2015 Fast r-CNN *Proc. IEEE Int. Conf. Comput. Vis.* p 1440
- [42] Arjovsky M, Chintala S and Bottou L 2017 Wasserstein generative adversarial networks *Proc. Int. Conf. Mach. Learn.* **70** 214
- [43] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need *Proc. Adv. Neural Inf. Process. Syst.* **30** 5998
- [44] Hinton G *et al* 2012 Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups *IEEE Signal Proc. Mag.* **29** 82
- [45] Devlin J, Chang M W, Lee K and Toutanova K 2019 BERT: pre-training of deep bidirectional transformers for language understanding *Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* p 4171
- [46] Wu D, Liao M, Zhang W and Wang X YOLOP: you only look once for panoptic driving perception [arXiv:2108.11250](https://arxiv.org/abs/2108.11250)
- [47] Racah E, Beckham C, Maharaj T, Kahou S E and Pal C 2017 Extreme weather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events *Proc. Adv. Neural Inf. Process. Syst.* **30** 3402
- [48] Park J and Sandberg I W 1991 Universal approximation using radial-basis-function networks *Neural Comput.* **3** 2467
- [49] Hornik K 1991 Approximation capabilities of multilayer feedforward networks *Neural Netw.* **4** 251
- [50] Rackauckas C, Ma Y, Martensen J, Warner C, Zubov K, Supekar R, Skinner D, Ramadhan A and Edelman A 2001 Universal differential equations for scientific machine learning [arXiv:2001.04385](https://arxiv.org/abs/2001.04385)
- [51] Raissi M, Perdikaris P and Karniadakis G E 2019 Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations *J. Comput. Phys.* **378** 686
- [52] Karniadakis G E, Kevrekidis I G, Lu L, Perdikaris P, Wang S and Yang L 2021 Physics-informed machine learning *Nat. Rev. Phys.* **3** 422
- [53] Lu L, Meng X, Mao Z and Karniadakis G E 2021 DeepXDE: a deep learning library for solving differential equations *SIAM Rev.* **63** 2088
- [54] E W and Yu B 2018 The deep ritz method: a deep learning-based numerical algorithm for solving variational problems *Commun. Math. Stat.* **6** 1
- [55] Lu L, Jin P, Pang G, Zhang Z and Karniadakis G E 2021 Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators *Nat. Mach. Intell.* **3** 218
- [56] Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A and Anandkumar A Neural operator: graph kernel network for partial differential equations [arXiv:2003.03485](https://arxiv.org/abs/2003.03485)
- [57] Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A and Anandkumar A Fourier neural operator for parametric partial differential equations [arXiv:2010.08895](https://arxiv.org/abs/2010.08895)
- [58] Nelsen N H and Stuart A M The random feature model for input-output maps between banach spaces [arXiv:2005.10224](https://arxiv.org/abs/2005.10224)
- [59] Patel R G, Trask N A, Wood M A and Cyr E C 2021 A physics-informed operator regression framework for extracting data-driven continuum models *Comput. Meth. Appl. Mech. Eng.* **373** 113500
- [60] Lanthaler S, Mishra S and Karniadakis G E 2022 Error estimates for deepONets: a deep learning framework in infinite dimensions *Trans. Math. Appl.* **6** 1
- [61] Kovachki N, Lanthaler S and Mishra S 2021 On universal approximation and error bounds for Fourier neural operators *J. Mach. Learn. Res.* **22** 1–76
- [62] Pu J, Peng W and Chen Y 2021 The data-driven localized wave solutions of the derivative nonlinear Schrödinger equation by using improved PINN approach *Wave Motion* **107** 102823
- [63] Fang Y, Wu G, Wang Y and Dai C-Q 2021 Data-driven femtosecond optical soliton excitations and parameters discovery of the high-order NLSE using the PINN *Nonlinear Dyn.* **105** 603–16
- [64] Zhou Z and Yan Z 2021 Deep learning neural networks for the third-order nonlinear Schrödinger equation: solitons, breathers, and rogue waves *Commun. Theor. Phys.* **73** 105006
- [65] Wang L and Yan Z 2021 Data-driven rogue waves and parameter discovery in the defocusing nonlinear Schrödinger equation with a potential using the PINN deep learning *Phys. Lett. A* **404** 127408
- [66] Wang L and Yan Z 2021 Data-driven peakon and periodic peakon travelling wave solutions of some nonlinear dispersive equations via deep learning *Physica D* **428** 133037
- [67] Zhu W, Khademi W, Charalampidis E G and Kevrekidis P G 2022 Neural networks enforcing physical symmetries in nonlinear dynamical lattices: the case example of the Ablowitz-Ladik model *Physica D* **434** 133264
- [68] Li J, Chen J and Li B 2022 Gradient-optimized physics-informed neural networks (GOPINNs): a deep learning method for solving the complex modified KdV equation *Nonlinear Dyn.* **107** 781–92
- [69] Mo Y, Ling L and Zeng D 2022 Data-driven vector soliton solutions of coupled nonlinear Schrödinger equation using a deep learning algorithm *Phys. Lett. A* **421** 127739

- [70] Yin Z, Siahkoohi A, Louboutin M and Herrmann F J Learned coupled inversion for carbon sequestration monitoring and forecasting with Fourier neural operators arXiv:[2203.14396](#)
- [71] Wen G, Li Z, Azizzadenesheli K, Anandkumar A and Benson S M 2022 U-FNO-An enhanced Fourier neural operator-based deep-learning model for multiphase flow *Adv. Water Resour.* **163** [104180](#)
- [72] You H, Zhang Q, Ross C J, Lee C H and Yu Y Learning deep implicit Fourier neural operators (IFNOs) with applications to heterogeneous material modeling arXiv:[2203.08205](#)
- [73] Xu Z, Zhang J Q Y, Luo T, Xiao Y and Ma Z Frequency principle: fourier analysis sheds light on deep neural networks arXiv:[1901.06523](#)
- [74] Luo T, Ma Z, Xu Z-Q J and Zhang Y Theory of the frequency principle for general deep neural networks arXiv:[1906.09235](#)
- [75] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *Proc. 32th. Int. Conf. Machine Learn.* **vol 6** pp 448–56