

Learning S -matrix phases with neural operators

Vasilis Niarchos^{1,*} and Constantinos Papageorgakis^{2,†}

¹*ITCP and CCTP, Department of Physics, University of Crete, 71003 Heraklion, Greece*

²*Centre for Theoretical Physics, Department of Physics and Astronomy, Queen Mary University of London, London E1 4NS, United Kingdom*



(Received 21 May 2024; accepted 1 August 2024; published 23 August 2024)

We use Fourier neural operators (FNOs) to study the relation between the modulus and phase of amplitudes in $2 \rightarrow 2$ elastic scattering at fixed energies. Unlike previous approaches, we do not employ the integral relation imposed by unitarity, but instead train FNOs to discover it from many samples of amplitudes with finite partial wave expansions. When trained only on true samples, the FNO correctly predicts (unique or ambiguous) phases of amplitudes with infinite partial wave expansions. When also trained on false samples, it can rate the quality of its prediction by producing a true/false classifying index. We observe that the value of this index is strongly correlated with the violation of the unitarity constraint for the predicted phase and present examples where it delineates the boundary between allowed and disallowed profiles of the modulus. Our application of FNOs is unconventional: it involves a simultaneous regression-classification task and emphasizes the role of statistics in ensembles of neural operators. We comment on the merits and limitations of the approach and its potential as a new methodology in theoretical physics.

DOI: [10.1103/PhysRevD.110.045020](https://doi.org/10.1103/PhysRevD.110.045020)

I. INTRODUCTION

The vast majority of problems in physics and mathematics involve the study of different types of functional relations. On general terms, these relations can be viewed as maps between infinite-dimensional spaces of functions. Sometimes, the origin of these maps is well understood. For example, a function may be obtained as the solution to an integral or differential equation that involves other input functions (e.g. functions that specify the form of the equation, boundary conditions, etc.). Analytic solutions are usually tractable only in special cases, while generic situations are computationally hard and require approximate schemes and numerical methods.

There are also many contexts where the rules dictating the map of interest are either poorly understood or beyond the reach of the existing framework. This is common for interacting, nonperturbative quantum field theories (QFTs). For example, in QFTs with a standard Lagrangian formulation, one would like to understand the map between spacetime-dependent deformations of the action by arbitrary operators, expressed by source functions in spacetime,

and the partition function of the theory (or its functional derivative with respect to the sources). The partition function contains all the necessary information about the local correlation functions of the QFT, which are some of the main objects of interest in quantum physics. The traditional computation of the partition function goes through a path integral, which is typically difficult to evaluate and in many cases also difficult to properly define.¹

In recent years, investigations originating from string theory have also revealed many new examples of QFTs that do not seem to admit a Lagrangian formulation and therefore challenge the traditional Lagrangian and Hamiltonian framework of quantum theories. There is very little we can currently compute in such theories with existing methods. This fact has motivated a flurry of activity in the search for new nonperturbative approaches to QFTs. The modern conformal bootstrap and S -matrix bootstrap programs [2–4] are prominent examples.

For the above reasons, it is particularly interesting to develop novel methodologies that will allow us to better understand general maps between functions in various contexts. We are especially interested in situations where partial information from explicit solutions in special

*Contact author: niarchos@physics.uoc.gr

†Contact author: c.papageorgakis@qmul.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

¹Note that in the (super)gravity limit of the AdS/CFT correspondence [1] the map between sources and functional derivatives of the partition function reduces to the solution of partial differential equations in classical gravity with suitable boundary conditions. This translates the QFT problem back to the study of functional relations in the context of differential equations mentioned above.

tractable cases can be used to uncover hidden structures and achieve generalizations toward computationally hard generic regimes. Can data-driven methods help in this direction? Can they produce reliable results with quantifiable error and potentially new analytic understanding?

In this paper, we would like to probe these general questions in a very specific problem that concerns the relation between the modulus and the phase of scattering amplitudes in elastic $2 \rightarrow 2$ scattering at fixed energies. This relation, which is an important ingredient of S -matrix theory, is constrained by unitarity through a nontrivial integral equation [see Eq. (5) below]. Instead of solving this equation directly, we will attempt to rediscover it by “learning” it from the data of amplitudes with finite partial wave expansions, where both the modulus and phase are straightforward to compute as functions of real phase shifts.

We will study the relationship between modulus and phase (and the implications of unitarity) using a modern supervised machine learning technique: neural operators (NOs) [5,6]. Unlike standard neural networks that are good function approximators, neural operators are good approximators of maps between infinite-dimensional function spaces. Since we are seeking to learn the map between the modulus and phase of a scattering amplitude—both functions of the scattering angle—NOs present themselves as an appealing tool.

Our main goal in this context will be to explore:

- (i) to what extent NOs generalize knowledge from finite to infinite partial wave expansions, and
- (ii) how to quantify the reliability of the result assuming no prior knowledge of the unitarity constraint.

Toward that end, we will run a simultaneous “regression-classification” task by training the NOs on both true and false samples. Their output will contain an extra label, which will be called “fidelity index,” indicating whether the prediction should be kept as a reliable solution or get rejected. We will provide evidence that the fidelity index extracts nontrivial features of the true solutions and that its value correlates with the degree of violation of the unitarity equation.

Typically, NOs supplement other direct methods in the solution of complicated equations, commonly partial differential equations (PDEs). The above implementation of NOs in a simultaneous regression-classification task is unconventional; to the best of our knowledge similar applications have been thus far limited (for some recent studies of NOs in image classification see [7–10]).

The performance of a NO—and how it learns—for fixed hyperparameters and training datasets depends on various stochastic factors that play a role during the training process and are hard to quantify. We will therefore also propose that it is useful to study the *collective* behavior of NOs. In particular, we will present specific data exhibiting the improved properties of the “mean fidelity index.” We will argue that quantities like the mean fidelity index can be

useful and could play a role similar to the Martin parameter $\sin \mu$ [see Eq. (9)], which provides a partial characterization of the scattering amplitudes.

Setting NOs aside for a moment, another popular machine learning method that appears in the context of PDEs is physics informed neural networks (PINNs) [11]. In that case, neural networks are used to directly model the unknown function: the equation to be solved goes into the definition of the loss that the network tries to minimize during training. Recently, PINNs were used to directly solve for the unitarity equation, obtaining notable results [12]. We emphasize that our approach should be viewed as complementary with an orthogonal scope, because we are attempting to reconstruct the unitarity equation and its implications without using it directly.

The rest of this paper is organized as follows. We begin in Sec. II with an introduction of the physics problem and a summary of the key formulas and definitions used in the main text. In Sec. III we present the salient features of PINNs and NOs, along with useful references for the nonexpert reader. The main results of the paper appear in Sec. IV, which focuses on amplitudes with unique phases, and Sec. V, which discusses the subtle case of amplitudes with phase ambiguities. In both cases, we see that NOs can generalize nontrivially beyond their training set, learning important properties about the structure of the system. We elaborate on the efficiency, advantages, disadvantages and difficulties of the approach. We conclude in Sec. VI with a brief summary of our main results and a discussion of interesting future prospects.

II. BACKGROUND: MODULUS AND PHASE IN ELASTIC SCATTERING

The following discussion is restricted to elastic $2 \rightarrow 2$ scattering. In quantum scattering processes, we measure the differential cross section $\frac{d\sigma}{d\Omega}$, which is equal to the square of the modulus of the scattering amplitude $f(\theta)$,

$$\frac{d\sigma}{d\Omega} = |f(\theta)|^2. \quad (1)$$

The scattering amplitude, which is part of the asymptotic form of the wave function in nonrelativistic quantum mechanics, is a complex number

$$f(z) = b(z)e^{i\phi(z)}, \quad (2)$$

with modulus $b(z)$ and phase $\phi(z)$. We used $z := \cos \theta$ to express the dependence on the scattering angle θ . From the differential scattering cross section one reads off $b(z)$, but it is in principle difficult to extract the corresponding phase $\phi(z)$.

Mathematically, this task is easy when the scattering amplitude admits a *finite* partial wave expansion

$$f(z) = \frac{1}{k} \sum_{\ell=0}^L (2\ell+1) \sin \delta_{\ell} e^{i\delta_{\ell}} P_{\ell}(z) \quad (3)$$

in terms of $L+1$ phase shifts δ_{ℓ} . Both $b(z)$ and $\phi(z)$ are expressed in terms of the phase shifts. In this form, unitarity plays a simple role; it dictates that the phase shifts are real. In (3), k is the wave number of a nonrelativistic particle scattered by some potential in quantum mechanics and the $P_{\ell}(z)$ are Legendre polynomials.

A generic amplitude, however, admits an *infinite* partial wave expansion. At fixed energy (equivalently, fixed k) the rescaled amplitude $F(z) = kf(z)$ is an infinite superposition of partial waves

$$F(z) = B(z) e^{i\phi(z)} = \sum_{\ell=0}^{\infty} (2\ell+1) \sin \delta_{\ell} e^{i\delta_{\ell}} P_{\ell}(z). \quad (4)$$

In that case, finding the phase $\phi(z)$ for a given $B(z)$ is more complicated and the partial wave expansion is less useful.

Nevertheless, when formulated more generally, unitarity is a strong condition that nontrivially relates the modulus and the phase of a scattering amplitude. A standard argument² shows that unitarity imposes the integral constraint

$$\sin \phi(z) = \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 \frac{B(z_1)B(z_2)}{4\pi B(z)} \cos[\phi(z_1) - \phi(z_2)], \quad (5)$$

where

$$z_2(z, z_1, \phi_1) \equiv zz_1 + \sqrt{1-z^2} \sqrt{1-z_1^2} \cos \phi_1. \quad (6)$$

For a given $B(z)$, one would like to solve this equation to determine the corresponding phase $\phi(z)$.

In the existing literature, a significant amount of effort has been put into determining for which $B(z)$ there exist solutions for $\phi(z)$, either unique or multiple, and several associated bounds have been established. The so-called “dual bound” is derived by setting $z = 1$ in (5). This special case provides a necessary condition for $B(z)$ to be the valid modulus of a scattering amplitude,

$$\int_{-1}^1 dz_1 \frac{B(z_1)^2}{2B(1)} \leq 1. \quad (7)$$

Additional bounds on existence and uniqueness can be obtained by defining the function

$$K(z) := \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 \frac{B(z_1)B(z_2)}{4\pi B(z)} \quad (8)$$

²See [13,14] for a review of the argument in nonrelativistic quantum mechanics and [15,16] for a discussion in relativistic QFT. A related discussion also appears in [17,18].

and the “Martin parameter”

$$\sin \mu := \max_{-1 \leq z \leq 1} K(z). \quad (9)$$

For example, one can trivially show using (5) that

$$|\sin \phi(z)| \leq \sin \mu. \quad (10)$$

Moreover, it can be proven that, given a modulus $B(z)$, solutions for phases always exist when $\sin \mu \leq 1$ [19] but known arguments do not preclude the existence of solutions also for $\sin \mu > 1$. Polynomial (finite partial wave) amplitudes are unique if $\sin \mu \leq 1$ [19]. For amplitudes with an infinite number of partial waves the best bound on uniqueness is currently $\sin \mu < 0.86$ [20], but it is believed that phases should be unique up to $\sin \mu < 1$, [18,19].

There can also be multiple (ambiguous) phases corresponding to the same modulus, which do not include the trivial ambiguity where all the $\delta_{\ell} \rightarrow -\delta_{\ell}$ [and, therefore, $F(z) \rightarrow -F(z)^*$ via (4)]. For elastic scattering this degeneracy is twofold [21,22] and has been completely classified for finite partial waves with $L = 2, 3$ [23–25]. Phase ambiguities in $L = 4$ amplitudes have been discussed in [26]. Twofold ambiguous solutions can also be constructed for amplitudes with infinite partial wave expansions. It is interesting to ask what is the lowest possible value of $\sin \mu$ for the ambiguous solutions. For example, for $L = 2$ the lowest value of $\sin \mu$ is 2.6. An amplitude with the lowest known value of $\sin \mu \simeq 1.67$ was constructed recently using machine learning methods in [12].

III. PINNS, NEURAL OPERATORS AND PHYSICS INFORMED NEURAL OPERATOR

We next summarize some of the high-level features of PINNs and neural operators for the nonexpert reader and highlight their main differences.

Let us assume that we want to solve a system of equations for a set of unknown functions. In many applications, this is a system of partial differential equations, a system of integro-differential equations, or a set of algebraic equations. A natural machine learning approach is to use neural networks (NNs) as universal function approximators [27] to model the unknown functions and set up a training process where the parameters of the NNs are optimized to satisfy the prescribed system of equations with the least possible error.³ The domain of the functions is discretized on a collocation grid, and the corresponding

³This process involves the solution of a typically very high-dimensional nonlinear, nonconvex optimization problem with thousands, millions, or more, parameters. Stochastic gradient descent methods have proved very efficient in this context and algorithms like the adaptive momentum estimation (ADAM) [28] are popular choices.

error in the equations is evaluated and quantified in a scalar semipositive quantity, typically the mean squared error on the grid. This idea forms the basis behind PINNs [11] (related ideas go back to several papers from the 1990s, e.g. [29,30]) and constitutes an “unsupervised” approach: the algorithm generates its own data and tries to solve a problem associated with the specific system of equations. When the form of the equations changes (e.g. the source function in a PDE or the functions that describe the boundary/initial conditions), the PINN needs to be optimized from scratch.

Neural operators are another data-driven approach that employs NNs. In this case, the idea is to approximate the “solution operator” that maps the input functions (e.g. source functions, boundary/initial conditions) to the output functions solving the system of equations. To achieve this goal a NN with a more complicated architecture is employed. The latter is not merely the composition of linear operations and pointwise nonlinear actions of activation functions, but also convolutions that act nondiagonally on the domain of the input functions. Early discussions of neural operators (and related universal approximation theorems) also go back to the 1990s, e.g. [31]. In the present work, we will be employing a modern incarnation of the neural operator concept, the so-called Fourier neural operators (FNOs), which are constructed using convolution kernels defined in Fourier space [5,6]. Another approach that shares some common features with neural operators are the deep operator networks (DeepONets), [32]. We will not consider DeepONets in this paper.

The NO is a “supervised” machine learning method. The training is based on a dataset of ground-truth input-output pairs that teach the algorithm to map between the input and output function spaces. In typical applications, this dataset is generated by solving the system of equations of interest through some other method. It is worth noting that, although functions are defined on a grid during this process, NOs are discretization invariant and exhibit advanced performance in zero-shot superresolution—namely, they can be trained at low-resolution samples and compute at never-before-seen high resolutions [5,6]. Another obvious characteristic advantage of NOs is that, once trained, they can quickly find the solution for new inputs without further retraining, in contradistinction with PINNs. This is convenient if one scans over a landscape of input functions (as we will be doing later in this paper).

There is a plethora of applications of NOs to PDEs in the literature. A recent application of NOs to the time-dependent Schrödinger equation and scattering in non-relativistic quantum mechanics appeared in [33].

Recently, the authors of Ref. [12] employed the PINN approach to study the relation between the modulus and phase of the scattering amplitude in elastic $2 \rightarrow 2$ scattering, solving the unitarity equation (5). They produced remarkable results, including a new solution with

ambiguous phases that has the lowest known Martin parameter $\sin \mu \simeq 1.67$. This result improved the relevant bound for the first time in 50 years.

In this paper we do not want to simply repeat the analysis of [12] using NOs as an alternative machine learning method. For the reasons outlined in the Introduction, our main motivation is to explore to what extent we can learn the solutions together with the equation we are trying to solve. In the present work, that means learning the modulus/phase relation in Sec. II *without* using the unitarity equation (5). In this quest, we will be using the NOs in a rather unorthodox way. The NO will be trained on both true and false samples in a class of input functions where (5) is trivially satisfied and will be asked to uncover nontrivial structure underlying (5) outside this class rating its own performance and the quality of its predictions. We hope that this application will inspire other similar explorations in even more complicated problems, where the underlying equations are missing.

As a final comment, we would like to point out that it is also possible to combine the benefits of PINNs and NOs in a hybrid construction that trains NOs using the loss of the underlying equation like a PINN. This approach is called a physics informed neural operator (PINO) and has been explored in the context of PDEs in [34]. It would be interesting to explore potential improvements of the results in this work and [12] using PINOs.⁴

IV. UNIQUE PHASES

In this section, we train a NO on a set of random finite partial wave expansions with $L = 1, 2, 3$ to learn the mapping between the input modulus $B(z)$ and the output $\sin \phi(z)$, the sine of the corresponding amplitude phase. We assume that the relation is one-to-one and set up the training accordingly. Once trained, we test how well the NO predicts the phase of unseen amplitudes, e.g. amplitudes with an infinite partial wave expansion. We also explore ways to detect whether or not the prediction is reliable.

A. Neural operator setup I: Training on samples of valid solutions

We now present our first attempt at NO training. We begin by listing the hyperparameters used and detail the choice of training and test datasets, before testing for generalizations of the trained model. All the computations in this work were performed on NVIDIA A100 GPUs with 40 GB RAM.

⁴It would also be interesting to explore further related applications in the context of the S -matrix bootstrap, see, e.g. [35].

1. Hyperparameters and training

Hyperparameters. Using the Fourier neural operator implementation of [6], for which a well-explained documentation can be found on GitHub, we set up a 1D tensorized Fourier neural operator (TFNO) implemented in PyTorch with the following hyperparameters:

```

number of Fourier modes : n_modes = 50,
number of hidden channels : hidden_channels = 64,
number of projection channels : projection_channels = 512,
number of layers : n_layers = 4,
type of factorization : factorization = "tucker,"
rank : rank = 0.01.

```

This is a model with 76,849 parameters that are tuned during the training to produce an optimal NO. The training optimization was performed using ADAM [36] with learning rate 10^{-3} , weight decay 10^{-5} and batch size 256. Varying the above hyperparameters did not result in significant variations of the results.

Training. The training dataset is prepared in the following manner. We generate random samples of amplitudes with finite partial wave expansions

$$F(z) = B(z)e^{i\phi(z)} = \sum_{\ell=0}^L (2\ell+1) \sin \delta_{\ell} e^{i\delta_{\ell}} P_{\ell}(z) \quad (11)$$

sampling the random phase shifts δ_{ℓ} from a uniform distribution. 100,000 samples are collected for $L = 1, 2$, and 3, separately, providing a total of 300,000 amplitudes. For each of these amplitudes we read off their modulus $B(z)$ and the sine of their phase $\sin \phi(z)$. Afterward, we discretize $z = \cos \theta \in [-1, 1]$ on a uniform grid of 100 points⁵ to produce 300,000 100-dimensional vectors \vec{B} and 300,000 100-dimensional vectors $\vec{\sin \phi}$. The collection of \vec{B} vectors is converted to a PyTorch tensor that forms the input of the NO during training. Similarly, the collection of $\vec{\sin \phi}$ vectors is converted to a PyTorch tensor that forms the ground-truth output of the NO. We train on 98% of the samples (namely, 294,000 samples) and test on 2% (namely 6000 samples, evenly distributed across $L = 1, 2, 3$). The results reported below are based on a single training run of 6500 epochs.

We emphasize that once the trained NO has been obtained, it can be used to make very quick predictions

for any input modulus $B(z)$, in stark contrast with the PINN approach, where retraining from scratch is needed for every new input.

2. Tests against known results

Once the NO has trained on known samples, we investigate how well it generalizes both on the same class of data (training/test dataset), as well as on different classes of never-before-seen data. Predicting the phases of amplitudes in the latter case would indicate that the NO is able to learn the unitarity relation (5) and effectively solve it without having direct access to it.

Tests within the training-test dataset. For starters, we can ask about the quality of predictions inside the training/test dataset. In Fig. 1 we plot the ground truth (blue) and predictions (orange) of the trained NO for $\sin \phi(z)$ of three randomly chosen samples from the test dataset with $L = 1, 2, 3$, respectively. The plots of the ground truth and prediction are visually indistinguishable, indicating that the NO has trained well. To get a sense of the numerical size of the error in the plots of Fig. 1, for $L = 1$ the average relative error between the ground truth and prediction across the whole z grid is 0.4%. For $L = 2$ it is 1.1% and for $L = 3$ it is 0.9%. These numbers are typical in the test dataset. The percentage of samples that exhibit average relative error above 10% is 5.2%.

A first sample of tests on moduli with infinite partial wave expansions. A more interesting question concerns the extent to which the NO can generalize outside the training dataset. The first case we would like to discuss here are amplitudes with an *infinite* partial wave expansion. For concreteness, we will consider two examples of linear and quadratic moduli that were analyzed also in Ref. [12]. Later we will scan more extensively over the predictions for amplitudes with linear, quadratic, as well as cubic moduli.

The first example concerns amplitudes with linear modulus $B(z) = az + b$ (and $b > |a|$ for positivity). It is straightforward to check that these amplitudes do not have

⁵A remarkable feature of NOs is their capacity to efficiently implement zero-shot superresolution [37]. In the context of quantum $2 \rightarrow 2$ scattering, this gives us the ability to train on a grid of, say, 100 points and then easily make accurate predictions at higher resolutions. We did not see the need to go beyond the 100-point resolution in this problem, but it is good to keep in mind that this possibility exists.

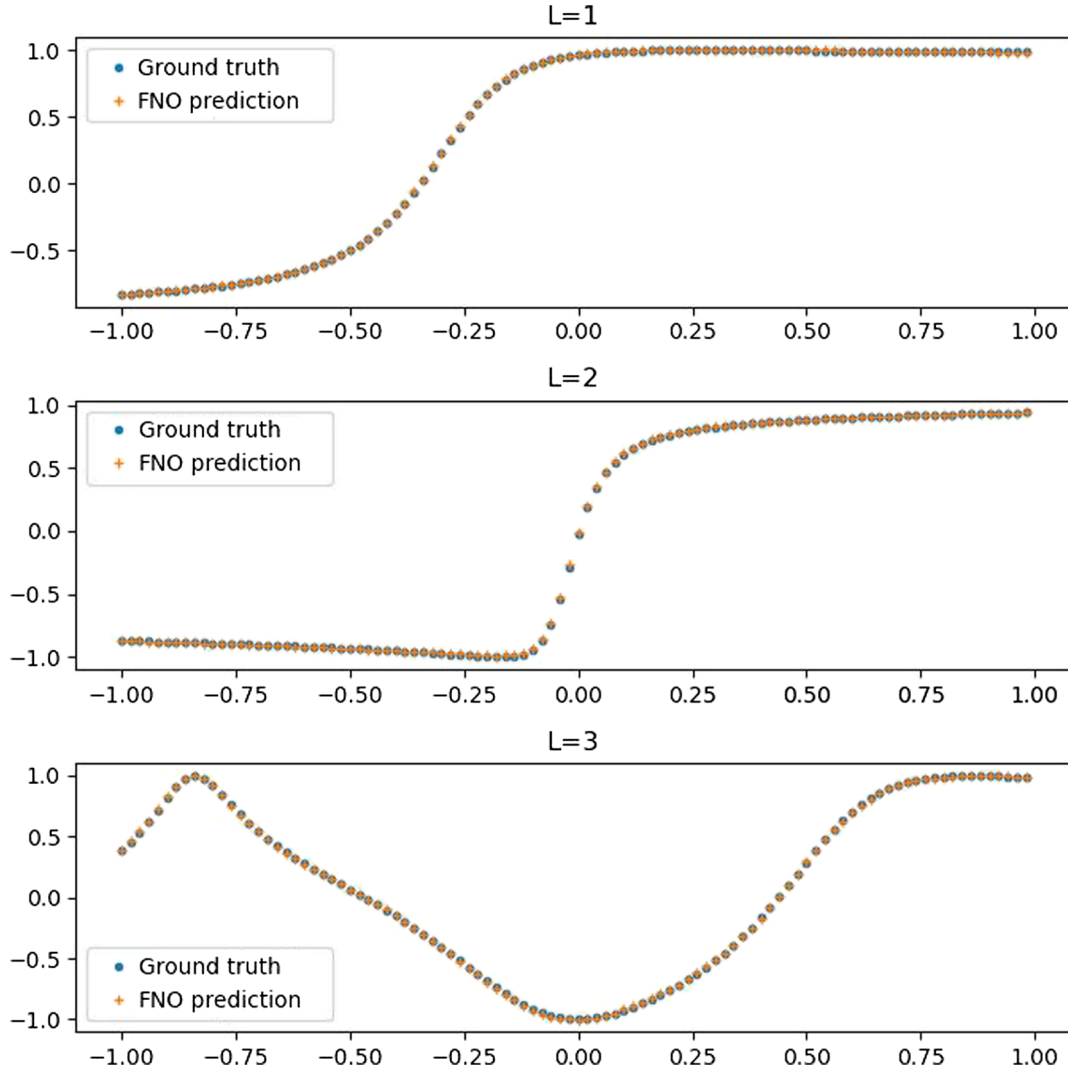


FIG. 1. Plots of the ground truth $\sin \phi(z)$ (blue color) and FNO-predicted $\sin \phi(z)$ (orange color) for three randomly chosen samples of amplitudes within the 6000 test dataset. From top to bottom we list plots for amplitudes with finite partial wave expansions and $L = 1, 2, 3$, respectively.

a finite partial wave expansion.⁶ In the top left plot of Fig. 2 we present the prediction of the NO for $\sin \phi(z)$ when $B(z) = \frac{1}{10}(z + 4)$ against a numerical solution of the unitarity equation (5) obtained with the use of an iteration scheme. This particular amplitude has Martin parameter $\sin \mu = 0.522$ and the iteration scheme converges very quickly. The NO prediction is denoted by orange, while the solution of the unitarity equation by blue. The two solutions are visibly close. On the top right plot of Fig. 2 we also present (point-by-point on our z grid) the relative difference $r(z)$ between the NO prediction $\sin \phi_{\text{NO}}(z)$ and the solution of the unitarity equation $\sin \phi_{(5)}(z)$,

$$r(z) := \left| \frac{\sin \phi_{(5)}(z) - \sin \phi_{\text{NO}}(z)}{\sin \phi_{(5)}(z)} \right|. \quad (12)$$

For most points the relative difference is of the order of $O(10^{-3})$.

The second example refers to the quadratic modulus $B(z) = \frac{1}{2}(z^2 + 1)$, which was also discussed in Ref. [12]. This amplitude has Martin parameter $\sin \mu = 0.867$ and can once again be determined numerically by solving the unitarity equation (5) with a simple iteration scheme. In the bottom left plot of Fig. 2 we present in orange and blue, respectively, the $\sin \phi$ for the NO prediction and the solution of the unitarity equation. Once again, the two plots are visibly close. In the bottom right plot of Fig. 2 we also present the relative difference, which is now of the order of $O(10^{-2})$ for most points. It increases near $z = -1$, where the prediction in the depicted run was less accurate.

⁶For a detailed discussion see [12].

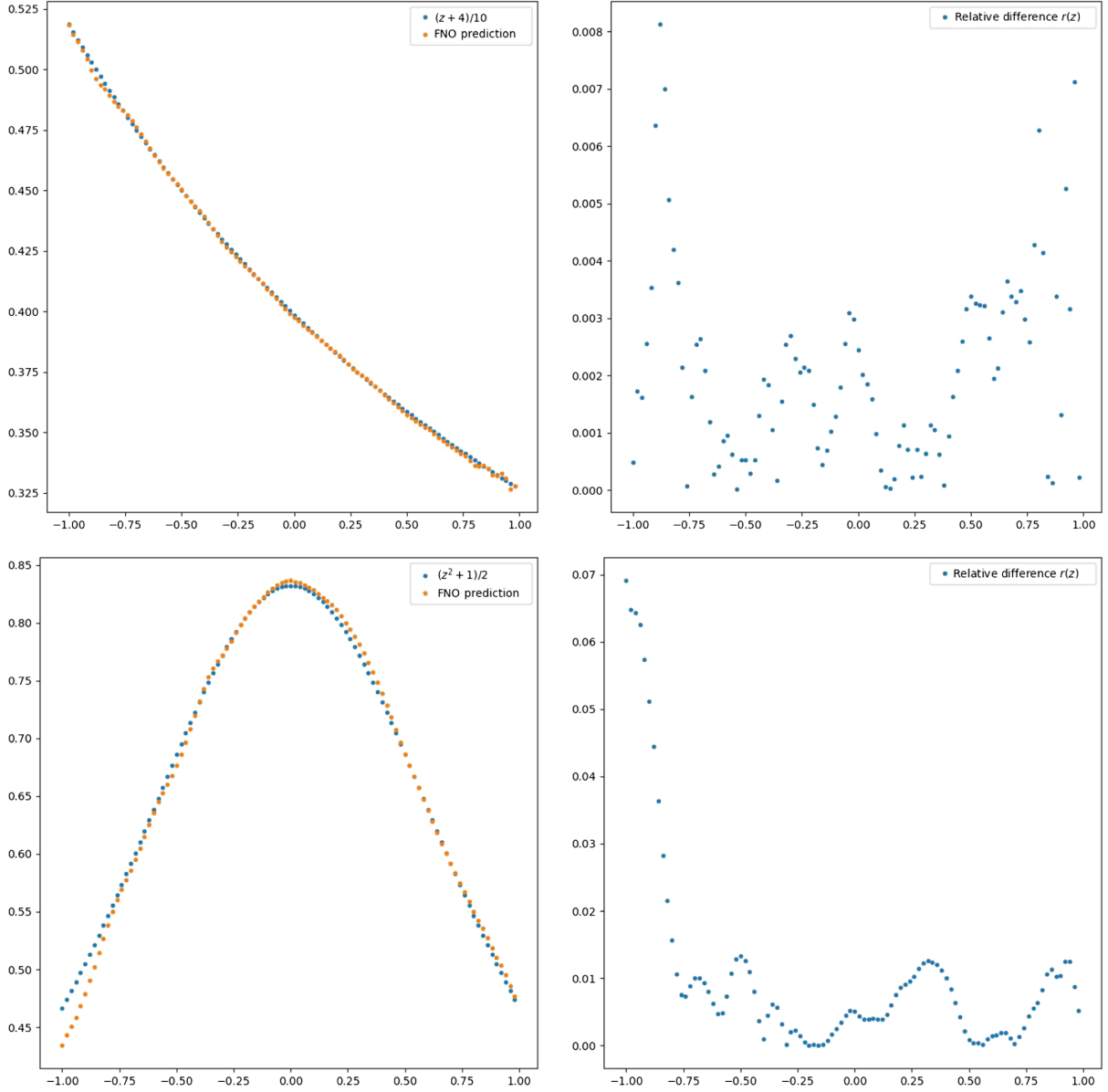


FIG. 2. The top two plots display the prediction of the trained NO for $\sin \phi(z)$ against the exact result for input modulus $B(z) = \frac{1}{10}(z+4)$. On the left are the actual functions, while on the right the pointwise relative difference. The bottom two plots display the corresponding data for input modulus $B(z) = \frac{1}{2}(z^2+1)$. Both cases refer to amplitudes with an infinite partial wave expansion.

Scans on linear, quadratic and cubic moduli. We can test the quality of the NO predictions on moduli with infinite partial wave expansions more extensively, by performing a scan over a wide grid of linear, quadratic and cubic moduli. To quantify the quality of the predictions we compute the loss in the unitarity condition (5),

$$\mathcal{L} := \frac{1}{N_c} \sum_{z_i} \left(\sin \phi(z_i) - \frac{1}{4\pi B(z_i)} \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) \times B(z_{2i}) \cos(\phi(z_1) - \phi(z_{2i})) \right)^2, \quad (13)$$

where z_{2i} is computed as in (6) with $z = z_i$. The sum is over the points z_i of the collocation z grid and the average is obtained by dividing with the number N_c of collocation points. In our runs $N_c = 100$. The integrals in (13) were computed numerically using the trapezoidal rule.⁷ Figure 3 displays the heat maps of the values of $\log_{10} \mathcal{L}$ for the NO

⁷The results presented in this section used the fixed grid of 100 collocation points in the NO training in order to apply the trapezoidal rule. It is straightforward to achieve higher numerical accuracy in the numerical computation of the integrals in (13) using higher resolution grids with NO zero-shot superresolution.

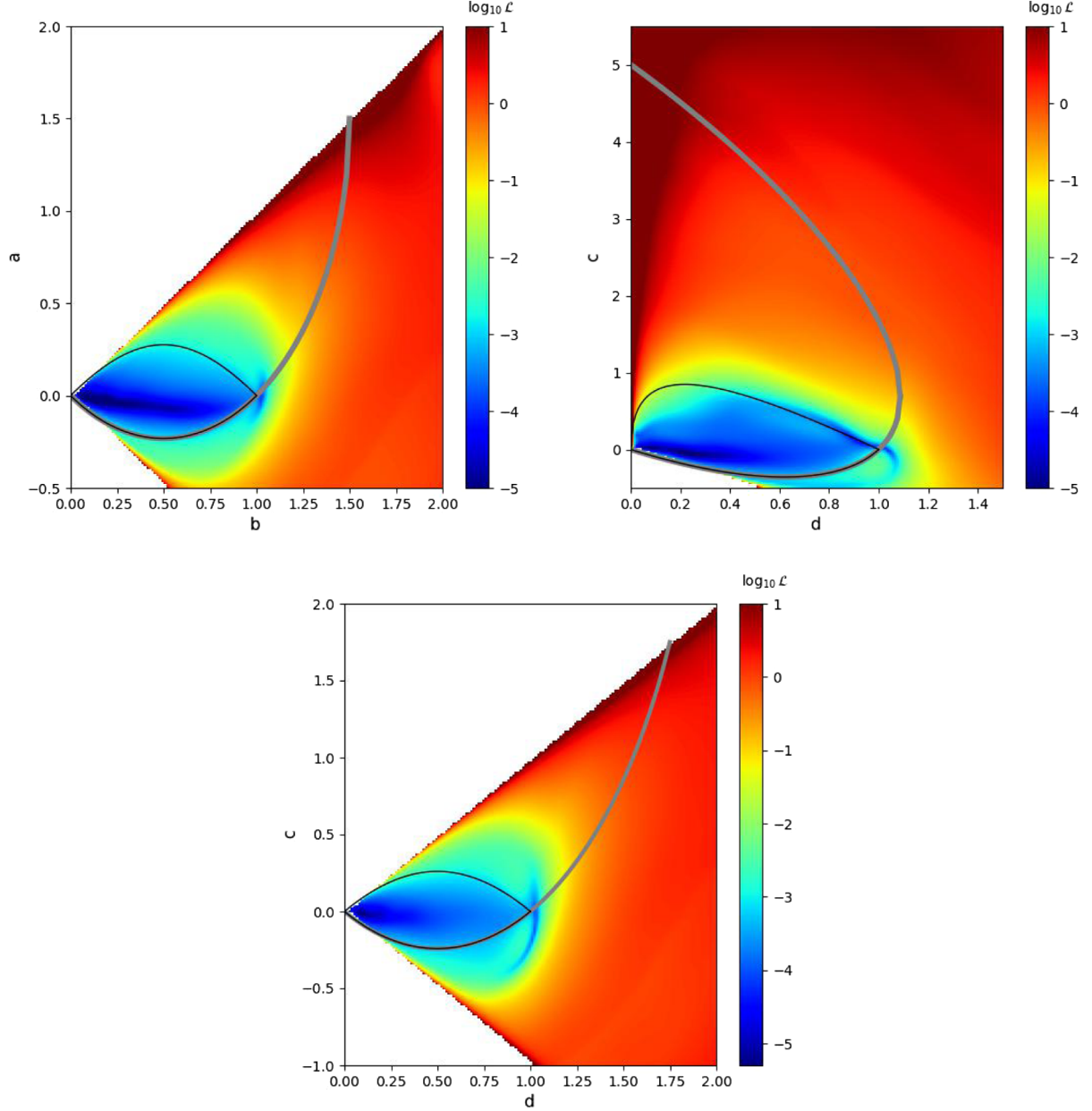


FIG. 3. Heat maps for the log base 10 loss of the NO prediction with respect to the unitarity condition (5). The top left plot refers to linear moduli $B(z) = az + b$, the top right plot to quadratic moduli $B(z) = cz^2 + d$ and the bottom plot to cubic moduli $B(z) = cz^3 + d$. Analogous results for the top two plots were obtained with the use of PINNs in Ref. [12] (see Figs. 3 and 5 of that paper). The thin black curves are the $\sin \mu = 1$ boundaries while the thick gray curves express the dual bounds.

predictions on a grid of linear, quadratic and cubic moduli. Let us comment on each of these plots separately.

For the linear moduli $B(z) = az + b$ we considered a grid of 180×150 points on the (a, b) plane for $a \in [-0.5, 2]$, $b \in [0, 2]$ and $b > |a|$. The heat map of the $\log_{10} \mathcal{L}$ values on this grid appears on the top left plot of Fig. 3. The corresponding heat map in Ref. [12] appears in Fig. 3 of that paper. Reference [12] computed on a grid of 75×60 points in the (a, b) plane, *retraining* the neural networks for 2K epochs to obtain each point. Instead, we are *evaluating* the already trained NO at each point

producing a heat map on a finer grid within approximately 20 sec.

In the top left plot of Fig. 3 we observe a distinct, blue-colored, low-loss region inside the $\sin \mu = 1$ contour, precisely like the one detected by PINNs for linear moduli in [12]. The main difference with the PINN result is that its lowest \log_{10} losses are in the vicinity of -8 , whereas our corresponding values are in the vicinity of -5 . That amounts to a difference in loss between the two methods at the level of 3 orders of magnitude. This is expected, since the PINN performs a dedicated optimization search for

each input modulus $B(z)$ explicitly using the unitarity condition (5), whereas the NO trains on a completely different class of inputs to produce a prediction outside its training dataset without using the unitarity condition. In that sense, the NO results in Fig. 3 are impressive and provide a distinct indication that the NO has been able to generalize well within the infinite partial wave amplitudes with linear moduli. Of course, by simply looking at the heat map of Fig. 3 one cannot really deduce where one should put the cutoff that separates the predictions that are consistent with unitarity from the ones that are inconsistent with unitarity. The same issue also exists within the PINN approach, but there it is slightly mitigated by the lower losses of the corresponding results. We will have to say more about how to address this difficulty in the next subsection.

For the quadratic moduli $B(z) = cz^2 + d$ we also considered a grid of 180×150 points on the (c, d) plane for $c \in [-0.5, 5.5]$, $d \in [0, 1.5]$ and $c > |d|$. The $\log_{10} \mathcal{L}$ heat map on this grid is depicted on the top right plot of Fig. 3. The corresponding heat map from Ref. [12] appears in Fig. 5 of that paper. Once again, we observe the formation of a low-loss region inside the $\sin \mu = 1$ contour, which is comparable with the result in Fig. 5 of Ref. [12], suggesting that the NO has been able to generalize to this class of amplitudes as well. Similar to the linear case, the NO losses are higher by roughly 3 orders of magnitude compared to the PINN losses of [12].

Finally, in the bottom plot of Fig. 3 we present the \log_{10} unitarity loss for cubic moduli of the form $B(z) = cz^3 + d$. Such amplitudes were not discussed in Ref. [12]. The resulting heat map is comparable to the linear-moduli heat map, exhibiting a distinct low-loss region inside the $\sin \mu = 1$ contour, as expected.

To summarize, in all three cases of infinite partial wave amplitudes analyzed in this subsection, the picture that emerges is impressively consistent with expectations from the analysis of the unitarity equation (5), suggesting that the NO has learned nontrivial features of that equation without having access to it. The results also exhibit some of the weaknesses of the approach:

- (a) The lowest losses are a few orders of magnitude higher than those produced by PINNs. That makes it harder to detect, without prior knowledge, the boundary between valid predictions consistent with unitarity and invalid predictions inconsistent with unitarity.
- (b) The low-loss regions (in blue) are not perfectly aligned with the $\sin \mu = 1$ and dual bounds. For example, there are small blue regions violating the dual bounds.
- (c) The quadratic-modulus NO heat map in Fig. 3 does not appear to detect the additional solutions appearing in Fig. 5 of [12] (e.g. the two small islands of $L = 2$ finite partial wave solutions).
- (d) To demonstrate nontrivial learning in the above cases, we had to use the unitarity relation (5). Without

explicit knowledge of that equation the heat maps in Fig. 3 would not have been possible. In addition, it is unclear to what degree of generality the NO has been able to learn the unitarity equation and whether it can make equally accurate predictions in arbitrary classes of infinite partial wave amplitudes.

We will return to these issues in the next subsection.

Tests on higher- L finite partial wave amplitudes. Another class of amplitudes outside the training-test dataset are the finite partial wave amplitudes with values of $L > 3$. Exploring the quality of the NO predictions in this class shows that already at $L = 4$ the NO fails to make any accurate predictions. This is, for example, apparent in the predictions presented at the bottom plot of Fig. 6, which depicts as a thick gray curve the exact result and as blue and orange dots the predictions of two separately trained NOs. The corresponding input modulus $B(z)$ appears on the top left plot of Fig. 6.

This case demonstrates that with the above-mentioned training on $L = 1, 2, 3$ amplitudes the NO cannot fully reconstruct the unitarity equation, which would allow for valid predictions with arbitrary input modulus $B(z)$. It has been able to learn nontrivial elements of the unitarity constraints, but not all the information that these constraints entail.

B. Neural operator setup II: Learning false predictions

The above observations raise the following related questions:

- (a) Can NOs learn to rate the quality of their predictions producing reliable results without any reference to the unitarity equation (5)?
- (b) Can NOs distinguish between moduli $B(z)$ that are allowed by unitarity and moduli that are not?
- (c) Can NOs uncover quantifiable elements of the unitarity equation without having access to it?

In this subsection we want to focus exclusively on results that can be obtained without any use of the unitarity equation. This immediately removes PINNs as a viable methodology. In general, asking whether we can obtain any results without the underlying equation is interesting, because there is a plethora of problems in physics and mathematics where knowledge of the underlying structure is missing.

In the setup of Sec. IV A, the NOs are designed to make a prediction for arbitrary input modulus $B(z)$. Without prior knowledge of the unitarity equation (5) it is impossible to deduce whether the solution exists, whether a prediction is valid, or to rate the quality of a prediction for a solution that exists. To address this difficulty, we propose setting up a slight variant of the NO of Sec. IV A, where the output has two components: the predicted $\sin \phi(z)$ and a classifying label that we call fidelity index \mathcal{F} , which contains information about the validity of the prediction. Accordingly, we

now train the NO on two types of $(\vec{B}, \overrightarrow{\sin \phi})$ samples: the first contains the moduli and phases of valid finite partial wave amplitudes and the second false moduli and phases that do not correspond to any amplitude. This setup should allow the NO to learn what it means to make a right prediction.

1. Hyperparameters and training

Hyperparameters. The results presented in this section were obtained with a 1D TFNO that has the same neural network and optimization hyperparameters as the model in Sec. IV A 1. However, in this case a different tensorization approach yields a larger model with a total number of 874,241 parameters (an order of magnitude larger than the one in the previous model of Sec. IV A 1).

Training. We are using the same z grid as in Sec. IV A 1 with 100 collocation points. The output vector of the NO is therefore 101 dimensional, including an extra element v_{101} characterizing the validity of the output. In our runs we chose to train by assigning the value 10 to valid input-output pairs and -10 to false pairs. The fidelity index was defined as $\mathcal{F} := \frac{10+v_{101}}{20}$, which assigns 1 to valid pairs and 0 to invalid ones.

We explored the results of training for a variety of datasets with varying fractions of true and false inputs/outputs. As one would expect, we observed that the quality of the classification output decreased when the fraction of false pairs was reduced. Here, we report results for a training-test dataset of 400,000 samples with the following composition: 75,000 true pairs for each of the $L = 1, 2, 3$ amplitudes and 175,000 false pairs. This yields a 43.75% fraction of false samples. The inputs and outputs of the false samples were generated randomly from two different groups of $L = 3$ amplitudes. We reserved 1200 samples for testing and the samples were randomly mixed to put the true and false pairs in random order. With these specifications, we trained 56 independent NOs for 1500 epochs.

2. Tests and observations

Tests within the training-test dataset. The accuracy of the fidelity-index prediction can be probed by computing the difference

$$\Delta\mathcal{F}(\text{sample}) := |\mathcal{F}_{\text{pred}}(\text{sample}) - \mathcal{F}_{\text{ground truth}}(\text{sample})| \quad (14)$$

between the predicted fidelity index and its ground truth, for each sample in the test dataset of 1200 samples and separately for each of the above 56 trained NOs. Assuming that a prediction is considered correct when

$$\Delta\mathcal{F} < C \quad (15)$$

for some arbitrarily chosen $C < 1$, we can go through the samples and register the number of times the inequality (15) is satisfied. This produces a success ratio $S_i(C)$ for the i th NO. We can further average this success ratio over the NOs; we call the corresponding quantity $\bar{S}_{\mathcal{F}}(C)$. For $C = 0.2$ and 0.3 we find

$$\bar{S}_{\mathcal{F}}(0.2) \simeq 73.94\%, \quad \bar{S}_{\mathcal{F}}(0.3) \simeq 75.34\%. \quad (16)$$

The values $S_{\mathcal{F}}^{(i)}(0.2)$ and $S_{\mathcal{F}}^{(i)}(0.3)$ for the individual NOs are very close to the above averages. In other words, there is little variation between the different NOs in this datum. This suggests that (with the above cutoffs for $\Delta\mathcal{F}$) the fidelity index makes the right classification roughly 75% of the time, which is an encouraging sign of classification capacity but not an impressively high percentage.

We can also rephrase the above test in terms of a mean fidelity index $\bar{\mathcal{F}}(\text{sample})$, which is defined for each sample as the average over the corresponding fidelity indices of all trained NOs. We can then define the difference $\Delta\bar{\mathcal{F}}(\text{sample})$ as in (14) by replacing \mathcal{F} with $\bar{\mathcal{F}}$, placing a cutoff as in (15) and computing the average $\bar{S}_{\bar{\mathcal{F}}}(C)$ over the samples. For this quantity we find

$$\bar{S}_{\bar{\mathcal{F}}}(0.2) \simeq 67.42\%, \quad \bar{S}_{\bar{\mathcal{F}}}(0.3) \simeq 73\%, \quad (17)$$

which is comparable to the previous result. We conclude that the average of the fidelity index over the NOs did not improve the classification capacity in this context.

These observations provide useful information about the performance of the NO as a classifier, but do not tell the whole story. In particular, we would now like to argue that the above tests do not really address some important aspects of the classification performance. Indeed, when one uses a NO to make a prediction for a never-before-seen modulus, it is very useful to know whether a predicted phase truly exists and can be considered correct with confidence, given an appropriately high fidelity index. Everything outside a small range of high fidelity values near 1 can be considered either as plausibly false for an existing phase or false because a phase does not exist. This viewpoint rephrases the way we should measure the success ratio in the test dataset.

Accordingly, we can now perform the following test. For each individually trained NO, we scan through the test dataset and count how many times the NO falsely affirms that the prediction is correct. The criterion for a prediction to be declared correct is

$$|\mathcal{F}_{\text{pred}}(\text{sample}) - 1| < C. \quad (18)$$

As we scan through the samples we count the cases where this inequality is satisfied and the ground-truth fidelity index vanishes (namely, the sample is false). That gives a

percentage⁸ of failure $f^{(i)}(C)$ for the i th NO. We want to examine if we can choose a small enough C in (18) that yields high confidence in true predictions [that is, small $f^{(i)}(C)$], but we also want to check how many true cases we missed with this criterion. We can also compute the average over the NOs

$$\bar{f}_{\mathcal{F}}(C) = \frac{1}{N_{\text{ops}}} \sum_i f^{(i)}(C), \quad (19)$$

where $N_{\text{ops}} = 56$ is the number of NOs. The label \mathcal{F} in $\bar{f}_{\mathcal{F}}(C)$ is there to remind us that we are using the fidelity index of the individual NOs to evaluate the criterion (18) (this will change in a moment). As above, we did not observe significant variation in $f^{(i)}(C)$ among different NOs. Therefore, we quote here the values of $\bar{f}_{\mathcal{F}}(C)$ for $C = 0.01, 0.02, 0.05, 0.1, 0.2$ along with the fraction of correct predictions of true samples over the total number of true samples

$$\begin{aligned} \bar{f}_{\mathcal{F}}(0.01) &= 4.08\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 66.5\% \\ \bar{f}_{\mathcal{F}}(0.02) &= 5.37\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 67.3\% \\ \bar{f}_{\mathcal{F}}(0.05) &= 8.00\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 68.3\% \\ \bar{f}_{\mathcal{F}}(0.10) &= 9.99\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 69.2\% \\ \bar{f}_{\mathcal{F}}(0.20) &= 11.71\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 69.8\%. \end{aligned} \quad (20)$$

We notice that the predictions of a true solution are only 4.08% of the times wrong when the fidelity index is inside the interval $[0.99, 1.01]$. This implies relatively high confidence in such predictions. We also notice that this criterion captures 66.5% of the total number of true samples in the test dataset. As we increase C (and with it the corresponding range of accepted fidelity indices) the fraction of wrong predictions increases and our confidence goes down, but the fraction of correct true predictions saturates. This implies that a small value of C at the level of 0.01 is a preferable choice.

It is also interesting to reevaluate these numbers using the mean fidelity index $\bar{\mathcal{F}}$. In that case, we are first averaging the fidelity index over the trained NOs for a given sample to produce the corresponding mean fidelity index $\bar{\mathcal{F}}_{\text{pred}}(\text{sample})$, then we use it to impose a criterion

like (18) and accordingly count which of the allowed samples are false predictions. This procedure yields a percentage of failure $f_{\bar{\mathcal{F}}}(C)$ for the “mean NO” and the analog of (20) is

$$\begin{aligned} f_{\bar{\mathcal{F}}}(0.01) &= 1.42\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 61.6\% \\ f_{\bar{\mathcal{F}}}(0.02) &= 1.56\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 65.2\% \\ f_{\bar{\mathcal{F}}}(0.05) &= 1.97\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 66.4\% \\ f_{\bar{\mathcal{F}}}(0.10) &= 2.58\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 67.0\% \\ f_{\bar{\mathcal{F}}}(0.20) &= 4.63\%, & \frac{\text{correct true predictions}}{\text{total no. of true samples}} &= 67.1\%. \end{aligned} \quad (21)$$

We observe that the mean fidelity index produces a lower percentage of failure at the same value of C (compared to the index of individual NOs) and, therefore, can be used to make predictions of correct phases with greater confidence. For example, the percentage of failed true predictions for the mean fidelity index at $C = 0.01$ is only 1.42%, compared to 4.08% of the individual fidelity indices. The fraction of correct true predictions is comparable in both cases meaning that the mean fidelity index continues to detect essentially the same number of true samples with higher confidence.

Correlations between the unitarity loss and the fidelity index. As a further calibrating question we can ask whether the fidelity index correlates with the values of the unitarity loss. As an example, in Fig. 4 we contrast the heat map of the \log_{10} unitarity loss and the heat map of the fidelity index for the predictions of one of the 56 trained NOs on the quadratic moduli $B(z) = cz^2 + d$. There is visible correlation between the plots, and this is typical in many training runs both for linear and quadratic moduli. It is difficult, however, to make a precise quantitative statement about their relation.

We also notice a clean separation between the predictions with high fidelity index (above 0.99) and predictions with low fidelity index (below 0.95). This is an interesting feature that correlates well with the above-mentioned observations about the fidelity index and its success rate. Unlike the unitarity loss, which varies smoothly between true and false predictions, the fidelity index appears to provide a more sharp acceptance/rejection criterion.

The results of the 56 trained NOs warrant some additional observations. First, we notice that the presence of the extra label that classifies the sample as true or false has affected the nature of the predictions across the landscape of input moduli. This is visible in the comparison of the

⁸We define the ratio that gives this percentage as the number of false predictions satisfying (18) divided by the total number of predictions satisfying (18).

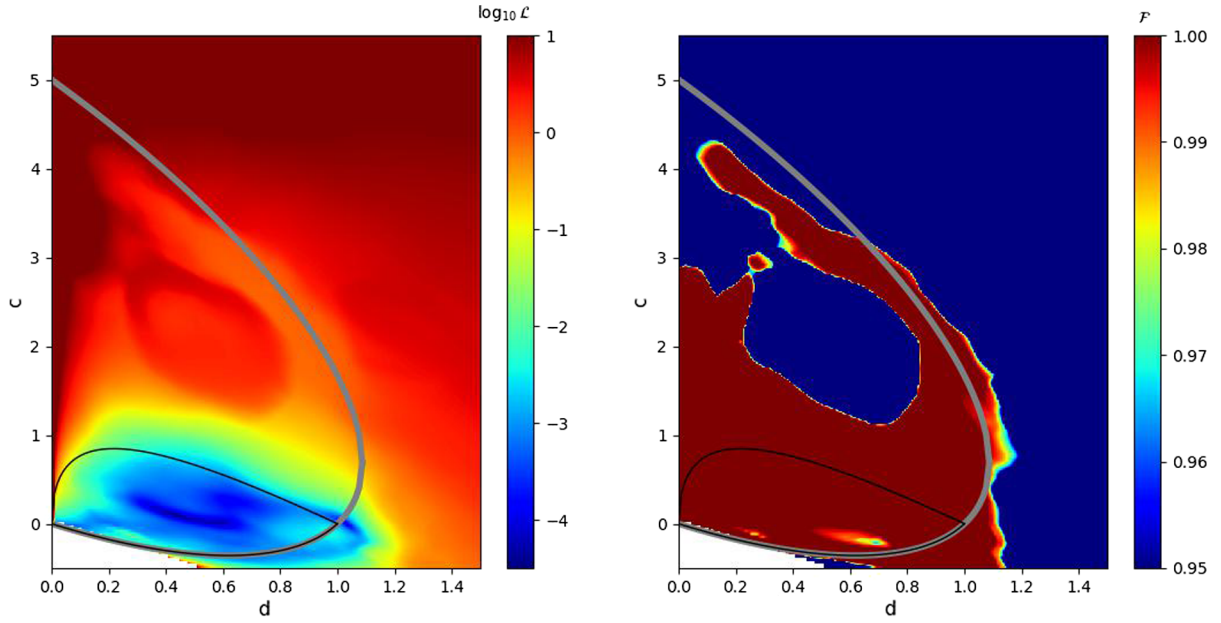


FIG. 4. Predictions for quadratic moduli $B(z) = cz^2 + d$ by one of the 56 NOs trained on both true and false S -matrix phases. The heat map on the left presents the \log_{10} unitarity loss of the predictions. The heat map on the right presents the value of the fidelity index \mathcal{F} . The color bar scale for the latter focuses on values between 0.95 and 1. Values below 0.95 are depicted in deep blue and values above 1 are depicted in deep red. As in previous plots, we have included the curve at $\sin \mu = 1$ (light black) and the curve of the dual bound (thick gray).

unitarity losses in the top right heat map of Fig. 3 against the heat map on the left of Fig. 4.

Second, the unitarity losses for the predictions of the 56 NOs that included the fidelity label are typically slightly higher than those for the predictions of the NOs in Sec. IVA 2, which did not involve any training on false samples. This is expected, since we were training with 300,000 true samples in Sec. IVA 2, whereas the training here involves a smaller number of true samples, 225,000.

Third, across the set of the 56 different NOs, we observed significant variation in the heat maps of the predicted phases and their fidelity index. This observation hints at the complexities of the training process in this context and makes it harder to extract invariant information from individual training runs. It is therefore interesting to explore whether we can obtain information independent of the fluctuations of individual training iterations, reflecting real properties of the system, by collecting statistics from multiple runs.

Performance of the mean fidelity index. The above discussion is further motivation in favor of the use of the mean fidelity index $\bar{\mathcal{F}}$, which involves an average over independent NOs. In what follows, we evaluate $\bar{\mathcal{F}}$ across the 56 previously trained NOs and plot the results in Fig. 5 across the landscape of linear and quadratic moduli. We observed that as we incorporated more and more NOs into the mean, there was an apparent convergence to the heat maps of Fig. 5. In the process, random fluctuating patterns from individual runs disappeared.

For both linear and quadratic moduli we notice that the averaging over NOs preserves the sharp transition between high and low fidelity indices that was characteristic in individual runs. Additionally, we once again observe that the high fidelity index regions in red (with values above 0.99) match well with the expectations from the test dataset for high confidence true predictions based on the mean fidelity index in this range.

For linear (quadratic) moduli, $\bar{\mathcal{F}}$ is plotted on the left (right) heat map of Fig. 5, together with the $\sin \mu = 1$ and dual bounds. We notice the characteristic concentration of high fidelity values around the $\sin \mu < 1$ region, which indicates that in the vicinity of this region the NOs correctly recognize predictions with the expected qualitative features of valid solutions.

The heat map of $\bar{\mathcal{F}}$ in the quadratic scan exhibits some additional intriguing features that seem to fit well with features of the heat map derived in Fig. 5 of Ref. [12] using PINNs. The high fidelity red region in our plot stretches above the upper $\sin \mu = 1$ boundary in a manner that seems to correlate with a region of relatively low-loss solutions [at the orders of $O(10^{-4.5})$ – $O(10^{-4})$] detected by PINNs.⁹ In addition, our heat map has a characteristic upward tail that trails closely the dual bound in the vicinity of two points:

⁹In the heat map of Fig. 5 in [12] the unitarity losses inside the $\sin \mu = 1$ contour are of the order of 10^{-8} . Here, we are referring to the losses of the PINN solutions right above that region and below the dual bound.

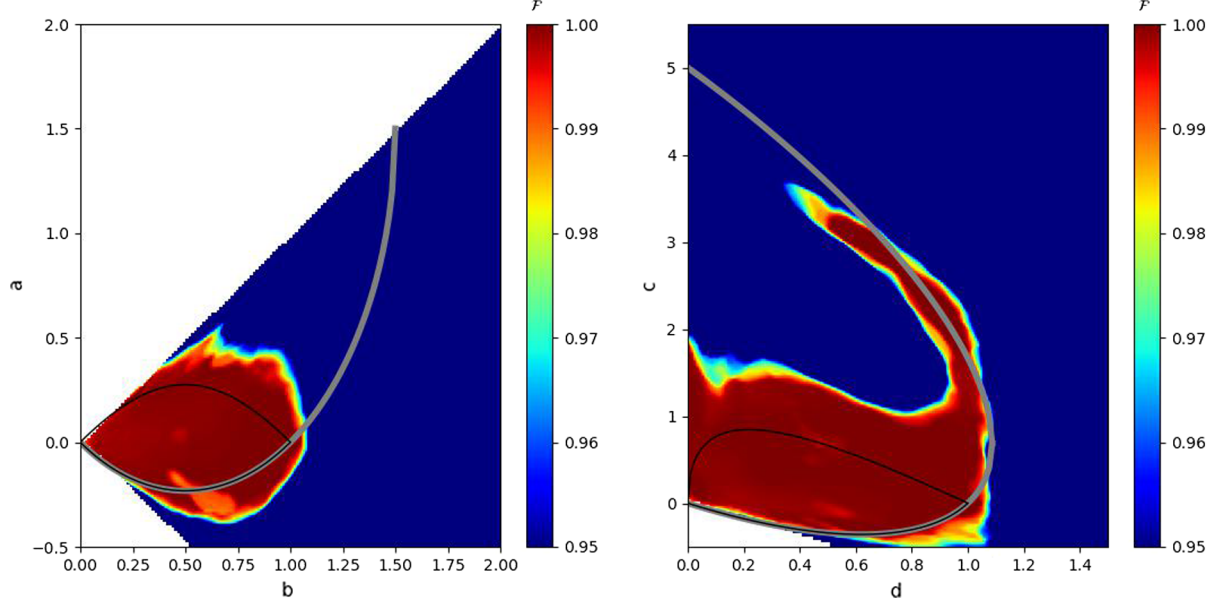


FIG. 5. The left heat map depicts the value of the mean fidelity index $\bar{\mathcal{F}}$ on the landscape of linear moduli $B(z) = az + b$. The right heat map depicts the corresponding values of the mean fidelity index for quadratic moduli $B(z) = cz^2 + d$. Notice that the color bar scale focuses on values between 0.95 and 1. We have also included the curve at $\sin \mu = 1$ (light black) and the curve of the dual bound (thick gray).

one at $(c, d) \simeq (3, 0.65)$ and another at $(c, d) \simeq (2.4, 0.9)$. The first point is tantalizingly close to the values $(c, d) = \sqrt{\frac{3}{8}}(5, 1) \sim (3.06, 0.61)$ of one of the finite partial wave solutions with $L = 2$ that the PINNs detect. The second point is similarly close to the values $(c, d) = \frac{5}{4}\sqrt{\frac{3}{7}}(3, 1) \sim (2.45, 0.82)$ of the second finite partial wave solution with $L = 2$ that the PINNs detect. Near the second point, our red region violates slightly the dual bound and so does a similar yellow blob in Fig. 5 from Ref. [12]. Interestingly, the NO of Sec. IV A 2, without the mean fidelity index, was unable to detect these features in the quadratic heat map of Fig. 3, but the NOs with the mean fidelity index seem to have picked them up. That is another indication that $\bar{\mathcal{F}}$ is a promising measure for the detection of real features learned by the NOs.

Detecting false predictions. We provided evidence that a high fidelity index (in the interval $[0.99, 1.01]$) can confidently assess that the prediction is a valid solution. Outside this interval the validity of the prediction is less clear, but it is natural to expect that a low fidelity index will be associated more frequently with a false prediction. We would next like to examine more closely how the fidelity index behaves in situations where the NO fails. For that purpose, we return to the finite partial wave amplitudes with $L = 4$, which proved to be a challenge for the NOs of Sec. IV A 2.

In Fig. 6 we plot the fidelity index and some of the actual predictions from the 56 NOs trained on a combination of true and false samples. The NOs have been evaluated on the

modulus of a randomly chosen $L = 4$ amplitude, which is depicted on the top left plot of Fig. 6. On the top right plot, we present the values of the fidelity index for each of the trained NOs. The vast majority of the NOs exhibit a low index with mean value 0.499; this is consistent with the fact that the NOs fail to correctly reproduce the corresponding phase, as one can explicitly check by plotting the predicted output for $\sin \phi(z)$ against the exact result. This is useful: the NOs fail to generalize in this case, but they recognize correctly that this is the case and provide a clear indication of that information in the output.

Looking closer at the fidelity indices for each of the 56 NOs in the top right plot of Fig. 6, we also notice that out of the 56 NOs only two have a fidelity index within the interval $[0.95, 1]$. They are denoted with a red circle in the top right plot of Fig. 6. These two points correspond to fidelity indices 0.957 and 0.962. According to the previous discussion, they lie outside the region that captures a confident true prediction. The predicted $\sin \phi(z)$ for these NOs are presented at the bottom plot of Fig. 6 against the exact result, denoted by the thick gray curve, in blue and orange, respectively. The orange prediction, which has the higher fidelity index 0.962, is clearly better and qualitatively closer to an acceptable solution. It is a smoother function within the interval $[-1, 1]$ (as one would expect from the sine of a real function), in contrast to the blue prediction that is chaotic and outside the interval $[-1, 1]$. The NO has recognized the importance of these features and has assigned a higher fidelity index to the orange prediction. In principle, it is impossible to exclude the orange prediction as false, but the fact that it stands as a

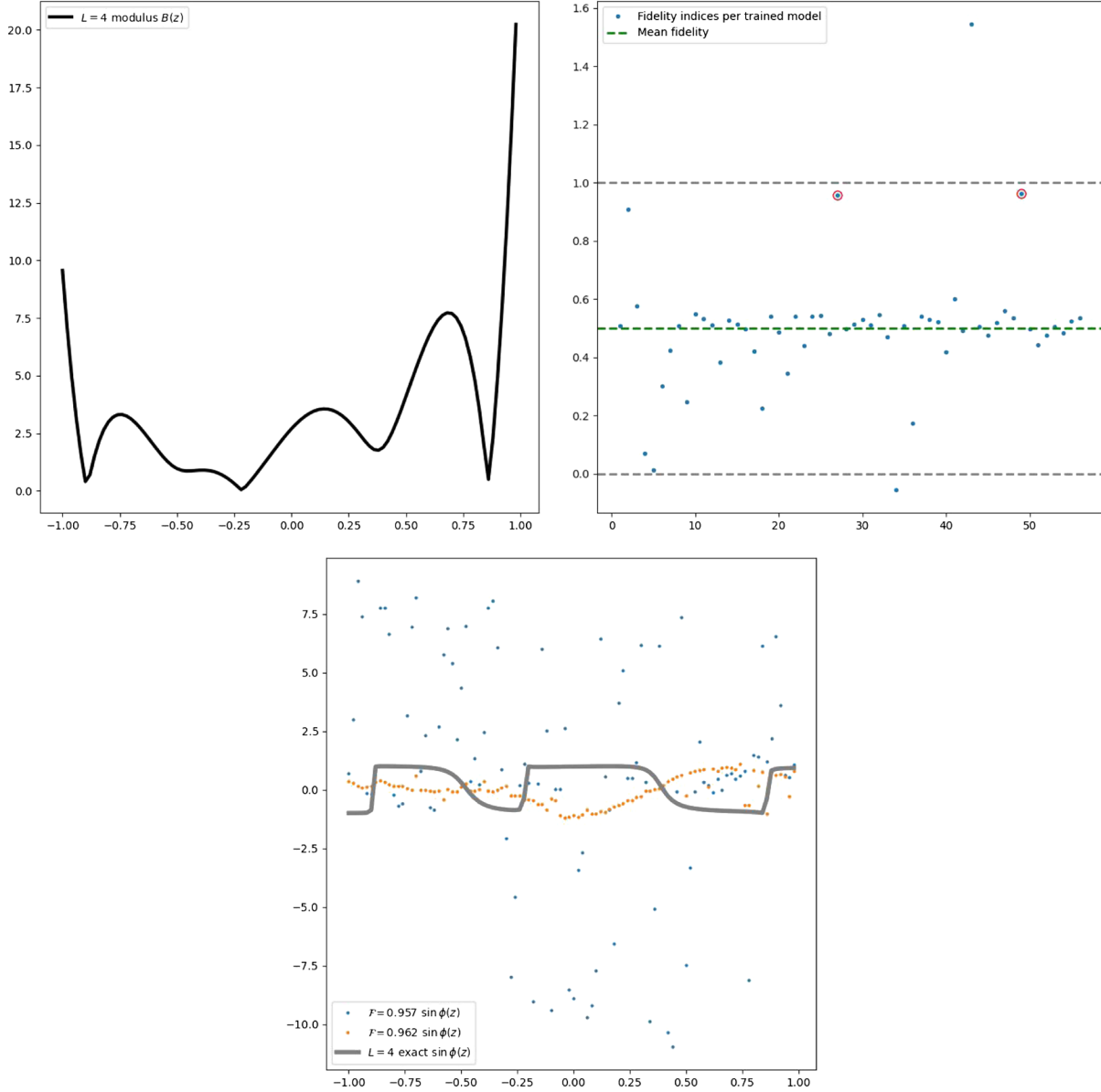


FIG. 6. The top left plot depicts the modulus of a random $L = 4$ amplitude. In the top right plot we present the fidelity indices for each of the 56 trained NOs evaluated on this specific modulus. The two points with a red circle around them represent the two predictions with fidelity indices closest to 1 (the actual values being 0.957 and 0.962, respectively). In the bottom plot we present the corresponding $\sin \phi(z)$ predictions of these two cases (in blue and orange) against the exact result represented by the thick gray curve.

clear outlier in the statistics of Fig. 6, and that the mean fidelity index is very low with small deviation, suggest that the orange prediction is likely false.

In conclusion, the above observations support using NOs within a statistical framework. In general situations, we propose the following approach: When the mean fidelity index suggests that a prediction should be rejected (as in Fig. 6), it should be discarded as potentially false. When the mean fidelity index is high (within the interval $[0.99, 1.01]$), one should accept the prediction as correct with high probability and extract predictions using the pointwise average of the predicted functions across the collection

of NOs. Useful information can also be extracted by the pointwise standard deviation of the predicted functions.¹⁰

V. AMBIGUOUS PHASES

The problem of S -matrix phases is interesting for an additional reason. So far we have operated under the

¹⁰A similar statistical approach was also advocated in the optimization schemes of Refs. [38,39]. In that context, the average of independent stochastic optimization runs (especially, those based on reinforcement learning) was always observed to provide better approximations.

assumption that for a given modulus $B(z)$ there is a unique solution for the phase $\phi(z)$ (up to trivial ambiguities) or that there is no solution at all. As we briefly reviewed in Sec. II, there are also cases of finely tuned moduli that admit a doubly ambiguous phase. Such cases were studied by several papers in the 1960s and 1970s and still lack a general complete classification. More recently, Ref. [12] revisited the construction of such solutions using the PINN approach. In this section, we would like to explore if we can detect the ambiguous solutions of infinite partial wave amplitudes by training NOs on unique and ambiguous solutions of finite partial wave amplitudes. For the training we are going to use the fully classified ambiguous amplitudes with $L = 2, 3$ [23–25]. Clearly, this task will be much more subtle and demanding, compared to the generic configurations we have been discussing so far.

A. Brief note on $L = 2, 3$ amplitudes with phase ambiguities

To generate training samples for $L = 2, 3$ amplitudes with ambiguous phases, we used the classification developed in [25]. Here, we briefly review the relevant construction and note some minor discrepancies in the original paper [25].

The approach of [25] involves an alternative decomposition of the partial wave amplitude in terms of the forward scattering amplitude (at $\theta = 0$) as

$$F(z) = F(1) \prod_{l=1}^L \frac{z - W_l}{1 - W_l}. \quad (22)$$

In this representation, all possible amplitudes with the same modulus at fixed L can be obtained by acting on (22) with the transformations [40]

$$\begin{aligned} S : \text{Re}F(1) &\rightarrow -\text{Re}F(1), \\ T_l : W_l &\rightarrow W_l^*. \end{aligned} \quad (23)$$

Combinations of the above symmetries are also allowed as long as they do not lead to phases that are trivially related by sending $\delta_\ell \rightarrow -\delta_\ell$. Defining the variables $\zeta_\ell := e^{2i\delta_\ell}$, it is straightforward to equate (11) with (22) and solve for $\zeta_\ell(W_\ell)$. One can then look for ambiguous solutions for the phase shifts δ_ℓ by requiring that (i) the $|\zeta_\ell|$ are left invariant by the transformations (23) and (ii) $|\zeta_\ell| = 1$, which is equivalent to imposing that the scattering is elastic. For $L = 2, 3$ this procedure leads to a real one-parameter family of twofold ambiguous phases (that are not trivially related) for specific intervals on the real line, as reported in Tables 1 and 2 of [25].

More specifically, for $L = 2$ the only independent transformation that does not lead to trivially related ambiguous phases is ST_1 . Following the above steps this recovers the

real one-parameter family of twofold ambiguous solutions of [24], including the Crichton ambiguity [23].

For $L = 3$ the only independent transformations that do not lead to trivially related ambiguous phases are T_1 and ST_1 . Analyzing the various possibilities leads to two classes of twofold ambiguous families of solutions arising for each of the T_1 and ST_1 transformations. In this context, we report the following disagreement with two of the expressions in [25]. We find

$$\cos \eta = \frac{\frac{15+3}{2x+7} + \frac{135|H'|^2x}{2} + \frac{8x}{15} + \frac{45}{2x} + 8}{3|H'|(4x+30)}, \quad (24)$$

$$|W_1|^2 = \frac{252c_5^2}{5c_7(7c_5^2 + 9c_1|A|^2)} - \frac{2c_1+1}{5c_5} + \frac{(2c_1+1)c_5}{c_7|A|^2} \quad (25)$$

for (A.5) and (A.14) in [25], respectively.¹¹ We are in agreement, however, with all other formulas, as well as the conclusions of the analysis of [25] as presented in their Tables 1 and 2.

B. Neural operators on the double cover

In order to incorporate the possibility of amplitudes that have the same modulus and two inequivalent phases, we set up a 1D TFNO that takes a single input $B(z)$, but outputs two $\sin \phi(z)$. On the z grid with 100 collocation points this implies that the output is a 200-dimensional vector, which concatenates the 100-dimensional vectors of the two predictions. When the prediction is unique, the concatenated vectors are identical. We will report results without a fidelity index, but that is a feature that can be readily incorporated in this discussion.

1. Hyperparameters and training

Hyperparameters. Following a simple grid search, we observed a significantly larger dependence of the results on the NO hyperparameters for this problem. In what follows, we will report results based on NOs with essentially the same hyperparameters as in Sec. IV A 1. The only hyperparameters that differ are the number of projection channels (we chose 256 instead of 512) and the number of layers (we chose 6 instead of 4). The resulting model has 72,745 parameters.

Training. We attempted training with several types of datasets involving different ratios of unique and ambiguous solutions. Since we were limited to a relatively small range of amplitudes with ambiguous phases at $L = 2, 3$, we could not significantly increase the total number of samples, which in turn made the training less efficient. The results presented below are based on a dataset with a total number of 100,000 randomly chosen samples and the following split: 30,000 random $L = 3$ amplitudes assumed to be

¹¹Note that in [25] what we call W_l is denoted as F_l .

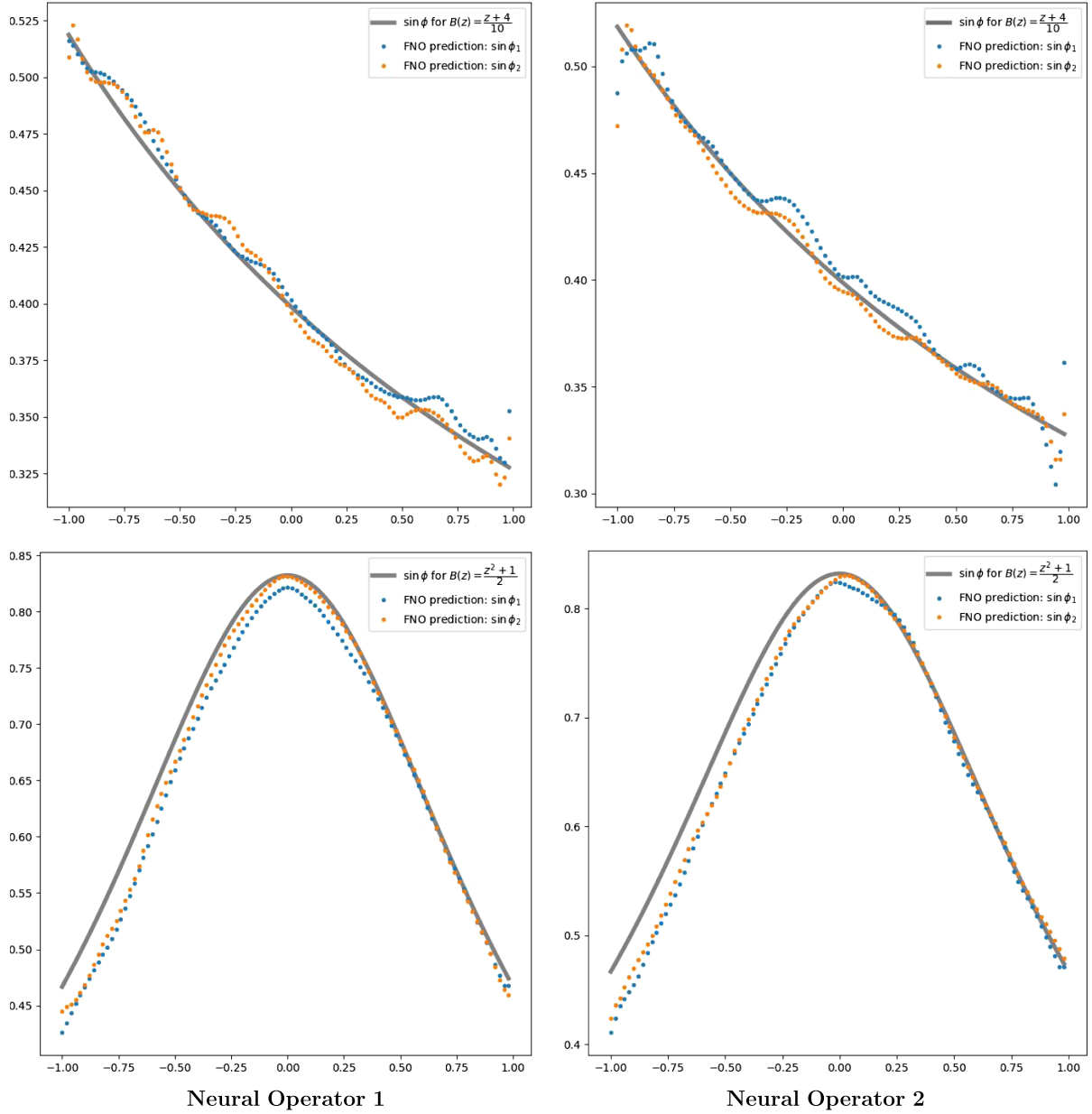


FIG. 7. The left column displays the exact $\sin \phi(z)$ (gray curve) and the two predictions of the first NO (blue and orange) for the linear and quadratic moduli $B(z) = \frac{1}{10}(z+4)$ and $B(z) = \frac{1}{2}(z^2+1)$. The right column displays the corresponding quantities for the second NO. Both NOs were trained on the same dataset and with the same hyperparameters.

unique, as well as 10,000 $L=2$ and 60,000 $L=3$ amplitudes with ambiguous phases sampled randomly across the different families of solutions summarized in the previous subsection. We trained on 99,000 of these samples and reserved 1000 samples for testing.

We present the results of two, independently trained, NOs with the same hyperparameters, which were trained for 6500 epochs.

2. Tests and observations

Once again, the NOs test well within the training-test dataset. Our purpose here is to explore whether they can

achieve any sensible generalization outside their immediate training domain. We will not attempt an exhaustive analysis, opting instead for the study of a few examples for illustration purposes. Specifically, we will focus on the performance of (a) the NOs on the linear and quadratic moduli of Fig. 2 [$B(z) = \frac{1}{10}(z+4)$ and $B(z) = \frac{1}{2}(z^2+1)$] that have an infinite partial wave expansion and no phase ambiguities and (b) one of the solutions with phase ambiguities in Ref. [41]—with parameter $z_1 = \frac{6}{5} + \frac{3}{5}i$ —that was also discussed in Ref. [12]; see e.g. Fig. 12 of that paper.

In Fig. 7 we present the predictions of the two NOs for the linear and quadratic moduli. In both cases, the two

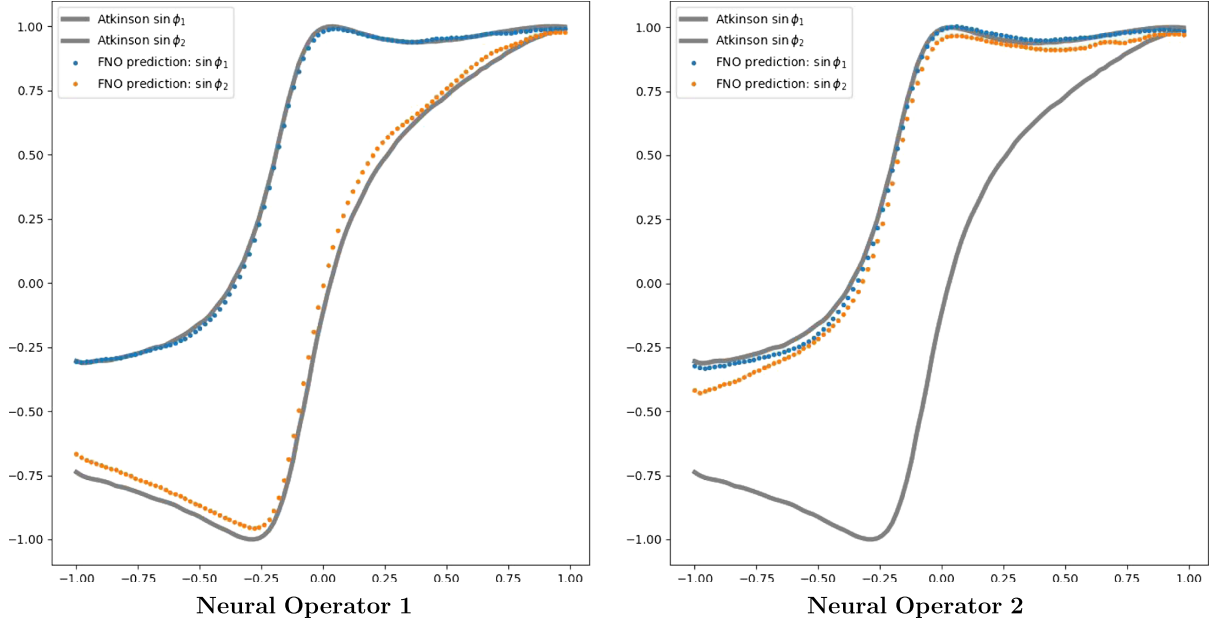


FIG. 8. The left plot displays the two predictions of the first NO (blue and orange) against the solutions for $\sin \phi(z)$ for the Atkinson *et al.* modulus with $z_1 = \frac{6}{5} + \frac{3}{5}i$. The right plot displays the predictions of the second NO. The first NO detected both solutions, while the second NO only one. Both NOs were trained on the same dataset, with the same hyperparameters and for the same number of epochs.

predicted phases are close to each other and close to the unique exact phase, but the accuracy of the results is obviously lower compared to the results of the previous sections. This is reasonable, since we only trained with 30,000 unique samples (compared to 300,000 unique samples in Sec. IV A 2).

In Fig. 8 we display the corresponding predictions of the two NOs for the Atkinson *et al.* [41] $z_1 = \frac{6}{5} + \frac{3}{5}i$ modulus. The first NO detects both phases, but the second detects only one of them. More generally, over several runs we observed that properly trained NOs would see either one or both solutions. More frequently, they would detect only one solution (the same one that the second NO detects in Fig. 8).

We also tested the above NOs on the $z_1 = 0.31 + 0.95i$ ambiguous amplitude of Ref. [12] that has $\sin \mu \simeq 1.67$; see Fig. 15 in that paper. The NOs predicted a unique output partially approximating one of the solutions of Ref. [12] with low accuracy. We observed that the prediction was more sensitive (compared to other inputs) to the precise numerics of the input modulus. This is an expected difficulty in general, as it involves generalization to measure-zero configurations and we have no dynamical way to tune the input modulus.

It would be interesting to explore if these problems can be addressed in the following manner: Still within the setting of not invoking the unitarity equation, one could first train a NO (or an ensemble of NOs) to produce a (mean) prediction of two $\sin \phi$ and a (mean) fidelity index. Then, in search of ambiguous solutions within the class where the NOs can generalize, one could run a PINN with a

NN that models the modulus $B_\theta(z)$ (with θ the NN parameters) and a loss function that has two contributions: (i) a repulsive potential for the two $\sin \phi$ and (ii) a potential involving the (mean) fidelity index, e.g. of the form $(\mathcal{F} - 1)^2$. Both contributions are functionals of $B_\theta(z)$ and the idea would be to optimize the PINN parameters θ so that it produces moduli with two inequivalent phases and a high fidelity index close to 1.

VI. CONCLUSIONS AND OUTLOOK

In this paper we used Fourier neural operators to study properties of amplitudes in elastic $2 \rightarrow 2$ scattering processes. Unlike previous approaches, we did not invoke the unitarity equation (5) to relate the modulus and phase, but tried to extract information about this relation from supervised training on random amplitudes with a finite partial wave expansion and $L = 1, 2, 3$. We observed that NOs can generalize nontrivially outside this class, successfully recovering (after a single training process) the heat maps of [12] for arbitrary linear and quadratic amplitude moduli. A similar approach was also applied to the twofold ambiguous phase solutions. Even though this case is generically much harder, as it concerns subtle properties of finely tuned configurations, it was nevertheless possible to demonstrate in specific examples that the NOs can generalize to recover two inequivalent phases for amplitudes with infinite partial wave expansions.

The question of how NOs generalize is not only central to this paper but also to the broader field of artificial intelligence (AI). The answer can depend on many factors,

which are usually hard to identify: the nature of the training dataset, the choice of hyperparameters and the details of the training, to name but a few. In the main text, we observed that within our specific setup the NO could learn several—but not all—nontrivial properties of the underlying general structure. For example, it could generalize to a class of amplitudes of infinite partial wave expansions, but failed on amplitudes with finite partial wave expansions for $L > 3$. Moreover, by simply training on true modulus-phase pairs, the NO could not detect the cases where a modulus is inadmissible. For that reason, it was crucial to train on both true and false samples, which were distinguished by an extra classifying label that we called the fidelity index. It was clear from several examples that this index could extract useful information about properties of scattering amplitudes, hidden inside the (inaccessible) unitarity equation (5). We emphasized the importance of averaging over independent NOs and provided evidence that it can be used to increase the confidence of the predictions and reduce optimization noise during training, enabling us to isolate true system information. In particular, the mean fidelity index made the predictions more robust and allowed the NO to rate its own performance.

We are excited by the potential use of similar approaches in other—possibly harder—problems, where the underlying structure is obscure, i.e. it is impossible to directly solve a system of equations or to directly compute relevant

quantities. For instance, it would be interesting to explore whether objects similar to the fidelity index can be defined (using NOs or other machine learning algorithms, especially generative AI algorithms) for other systems. In addition, the examples presented in this paper seem to indicate that by studying the statistics of learners for the same training dataset and hyperparameters, one can distill information about what this particular class of algorithms can—and cannot—learn without recourse to the unknown microscopics, hence providing a new road toward structures we do not yet understand.

ACKNOWLEDGMENTS

The work of V. N. was partially supported by the H.F.R.I. call “Basic research Financing (Horizontal support of all Sciences)” under the National Recovery and Resilience Plan “Greece 2.0” funded by the European Union—NextGenerationEU (H.F.R.I. Project No. 15384). The work of C. P. was partially supported by the Science and Technology Facilities Council (STFC) Consolidated Grants No. ST/T000686/1 and No. ST/X00063X/1 “Amplitudes, Strings & Duality.” Calculations were performed using the Sulis tier 2 HPC platform hosted by the Scientific Computing Research Technology Platform at the University of Warwick. Sulis is funded by EPSRC Grant No. EP/T022108/1 and the HPC Midlands+ consortium.

-
- [1] J. M. Maldacena, The large N limit of superconformal field theories and supergravity, *Adv. Theor. Math. Phys.* **2**, 231 (1998).
 - [2] D. Poland, S. Rychkov, and A. Vichi, The conformal bootstrap: Theory, numerical techniques, and applications, *Rev. Mod. Phys.* **91**, 015002 (2019).
 - [3] S. Rychkov and N. Su, New developments in the numerical conformal bootstrap, [arXiv:2311.15844](#).
 - [4] M. Kruczenski, J. Penedones, and B. C. van Rees, Snowmass white paper: S-matrix bootstrap, [arXiv:2203.02421](#).
 - [5] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, Neural operator: Learning maps between function spaces, *J. Mach. Learn. Res.* **24**, 4061 (2023).
 - [6] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, Fourier neural operator for parametric partial differential equations, [arXiv:2010.08895](#).
 - [7] W. Johnny, H. Brigido, M. Ladeira, and J. C. F. Souza, Fourier neural operator for image classification, in *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)* (2022), [10.23919/CISTI54924.2022.9820128](#).
 - [8] J. Xi, O. K. Ersoy, M. Cong, C. Zhao, W. Qu, and T. Wu, Wide and deep Fourier neural network for hyperspectral remote sensing image classification, *Remote Sens.* **14**, 2931 (2022).
 - [9] S. Kabri, T. Roith, D. Tenbrinck, and M. Burger, Resolution-invariant image classification based on Fourier neural operators, in *International Conference on Scale Space and Variational Methods in Computer Vision Lecture Notes in Computer Science* (Springer, Cham, 2023), p. 236.
 - [10] A. Kashefi and T. Mukerji, A novel Fourier neural operator framework for classification of multi-sized images: Application to three dimensional digital porous media, *Phys. Fluids* **36**, 057131 (2024).
 - [11] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* **378**, 686 (2019).
 - [12] A. Dersy, M. D. Schwartz, and A. Zhiboedov, Reconstructing S-matrix phases with machine learning, *J. High Energy Phys.* **05** (2024) 200.
 - [13] L. Landau and E. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Elsevier Science, New York, 1981).
 - [14] U. Buck, Inversion of molecular scattering data, *Rev. Mod. Phys.* **46**, 369 (1974).

- [15] A. Martin, *Scattering Theory: Unitarity, Analyticity and Crossing* Lecture Notes in Physics (Springer Berlin, Heidelberg, 2007).
- [16] M. Correia, A. Sever, and A. Zhiboedov, An analytical toolkit for the S -matrix bootstrap, *J. High Energy Phys.* **03** (2021) 013.
- [17] R. G. Newton, Determination of the amplitude from the differential cross section by unitarity, *J. Math. Phys. (N.Y.)* **9**, 2050 (1968).
- [18] D. Atkinson, Introduction to the use of non-linear techniques in s -matrix theory, *Acta Phys. Aust. Suppl.* **7**, 32 (1970).
- [19] A. Martin, Construction of the scattering amplitude from the differential cross-sections, *Nuovo Cimento A* **59**, 131 (1969).
- [20] A. D. Gangal and J. Kupsch, Determination of the scattering amplitude, *Commun. Math. Phys.* **93**, 333 (1984).
- [21] C. Itzykson and A. Martin, Phase-shift ambiguities for analytic amplitudes, *Nuovo Cimento A* **17**, 245 (1973).
- [22] A. Martin and J.-M. Richard, New result on phase shift analysis, *Phys. Rev. D* **101**, 094014 (2020).
- [23] J. H. Crichton, Phase-shift ambiguities for spin-independent scattering, *Il Nuovo Cimento A* (1965–1970) **45**, 256 (1966).
- [24] D. Atkinson, P. W. Johnson, N. Mehta, and M. De Roo, Crichton’s phase-shift ambiguity, *Nucl. Phys.* **B55**, 125 (1973).
- [25] F. A. Berends and S. N. M. Ruijsenaars, Examples of phase-shift ambiguities for spinless elastic scattering, *Nucl. Phys.* **B56**, 507 (1973).
- [26] H. Cornille and J. M. Drouffe, Phase-shift ambiguities for spinless and $4 \geq l(\max)$ elastic scattering, *Nuovo Cimento A* **20**, 401 (1974).
- [27] G. V. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.* **2**, 303 (1989).
- [28] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [29] T. Chen and H. Chen, Approximations of continuous functionals by neural networks with application to dynamic systems, *IEEE Trans. Neural Networks* **4**, 910 (1993).
- [30] I. E. Lagaris, A. Likas, and D. I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, *IEEE Trans. Neural Networks* **9**, 987 (1993).
- [31] T. Chen and H. Chen, Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems, *IEEE Trans. Neural Networks* **64**, 911 (1995).
- [32] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nat. Mach. Intell.* **3**, 218 (2019).
- [33] S. Mizera, Scattering with neural operators, *Phys. Rev. D* **108**, L101701 (2023).
- [34] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, Physics-informed neural operator for learning partial differential equations, [arXiv:2111.03794](https://arxiv.org/abs/2111.03794).
- [35] F. Bhat, D. Chowdhury, A. Sinha, S. Tiwari, and A. Zahed, Bootstrapping high-energy observables, *J. High Energy Phys.* **03** (2024) 157.
- [36] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [37] A. Shocher, N. Cohen, and M. Irani, “Zero-shot” super-resolution using deep internal learning, [arXiv:1712.06087](https://arxiv.org/abs/1712.06087).
- [38] G. Kántor, V. Niarchos, C. Papageorgakis, and P. Richmond, 6D (2,0) bootstrap with the soft-actor-critic algorithm, *Phys. Rev. D* **107**, 025005 (2023).
- [39] V. Niarchos, C. Papageorgakis, P. Richmond, A. G. Stapleton, and M. Woolley, Bootstrability in line-defect CFTs with improved truncation methods, *Phys. Rev. D* **108**, 105027 (2023).
- [40] A. Gersten, Ambiguities of complex phase-shift analysis, *Nucl. Phys.* **B12**, 537 (1969).
- [41] D. Atkinson, L. P. Kok, and M. de Roo, Crichton ambiguities with infinitely many partial waves, *Phys. Rev. D* **17**, 2492 (1978).