Machine learning a manifold

Craven, Sean

2022-05-25

Craven , S , Croon , D , Cutting , D & Houtz , R 2022 , ' Machine learning a manifold ' , Physical Review D , vol. 105 , no. 9 , 096030 . https://doi.org/10.1103/PhysRevD.105.096030

http://hdl.handle.net/10138/346187
https://doi.org/10.1103/PhysRevD.105.096030

Machine learning a manifold

Sean Craven,[1] Djuna Croon,[1,2,3,*] Daniel Cutting,[4,†] and Rachel Houtz[1,2,‡]

[1]*Department of Physics, Durham University, Durham DH1 3LE, United Kingdom*
[2]*Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, United Kingdom*
[3]*TRIUMF Theory Group, 4004 Wesbrook Mall, Vancouver, British Columbia V6T2A3, Canada*
[4]*Department of Physics and Helsinki Institute of Physics, PL 64, FI-00014 University of Helsinki, Finland*

(Received 5 January 2022; accepted 4 May 2022; published 25 May 2022)

We propose a simple method to identify a continuous Lie algebra symmetry in a dataset through regression by an artificial neural network. Our proposal takes advantage of the $\mathcal{O}(\epsilon^2)$ scaling of the output variable under infinitesimal symmetry transformations on the input variables. As symmetry transformations are generated post-training, the methodology does not rely on sampling of the full representation space or binning of the dataset, and the possibility of false identification is minimized. We demonstrate our method in the SU(3)-symmetric (non-) linear Σ model.

DOI: 10.1103/PhysRevD.105.096030

I. INTRODUCTION

Symmetry principles have drastically simplified the description of particle physics in the twentieth century. Famously, the 8-fold way [1] of organizing pions and kaons into a representation of an approximate SU(3) flavor symmetry lead to the development of the quark model. In the same vein, future discovery experiments would primarily have access to the low-energy particle content of theories beyond the standard model (BSM): in the case of a broken approximate global symmetry, this includes the pseudo-Nambu Goldstone bosons (pNGB), transforming under the adjoint representation of the unbroken symmetry (see, e.g., [2] for a relevant review.). If the BSM theory is confining, the symmetries of the low energy theory provide a window to the structure of the high energy theory through the barrier of the strong coupling regime. However, the pNGB representation need not have a small dimensionality, or define a simple topology. It may also be broken both spontaneously and explicitly, and the dataset may be noisy. Identifying residual (approximate) symmetries is therefore an interesting problem.

Motivated by this problem, we investigate the use of artificial neural networks (NN) to identify a symmetry in a dataset. We work with a simplified version of the problem: a function $V(\phi)$ symmetric under a transformation of coordinates $\phi \to f(\phi)$: $V(\phi) = V(f(\phi))$. To interpolate between datapoints we use a NN (recently discussed in the context of high energy physics in [3]), which allows us to test the local properties of the manifold and deduce the presence of a symmetry—or rather, eliminating the possibility of its absence—from its topology.

Detection of symmetry with the use of machine learning has a long history [4], though most attempts focus on mirror or rotational symmetries in image data and within the domain of computer vision [5–8]. In recent years there has been an increased interest in learning invariant transformations of input data which do not change the output of a specific machine learning task [9–16]. This is useful as the construction of invariant or equivariant NN reduces the number of samples of input data required for generalization. Machine learning has been used to explore various features of conformal field theories, including to distinguish between scale invariant and conformal symmetries [17]. It has also been demonstrated that computation of tensor products and branching rules of irreducible representations are machine-learnable [18]. Furthermore, a recent work has investigated using generative adversarial networks to learn transformations that preserve the measured probability density function of a random process [19]. Here we are interested in a variation to this problem, testing for the presence of a symmetry in a dataset that samples a patch of a function whose domain has a high dimensionality.

The use of NN for the detection of symmetries in such a context has previously been considered by [20,21] for translations, discrete symmetries, and $\mathrm{SO}(N) \simeq \mathrm{SU}(N-1)$ with $N < 3$. The methodology in this paper differs from the approaches taken in Refs. [20,21] in two important aspects. First, points related by a symmetry transformation are

*djuna.l.croon@durham.ac.uk
†daniel.cutting@helsinki.fi
‡rachel.houtz@durham.ac.uk

generated post-training. This implies that the local properties of the manifold can in principle be studied without global knowledge of the manifold, or a large number of close neighbors in the tangent space. Both of these implications may prove to be a marked advantage in datasets with large dimensionality. For example, no pretraining stage of narrow bin definition and data categorization is necessary. The fraction of data which is related by the action of a single generator roughly scales like $1-(\Delta y/y)^d$, where Δy is a narrow bin width in output variable y and where d is the dimensionality of the dataset. A further difference is that the methodology here can be used to demonstrate the absence of a symmetry, such that the probability of misidentification is minimized. With sparser sampling the assumptions made about the symmetry transformation (for example its direction) may play an increasingly important role, potentially leading to the false identification of SO(N) symmetry. We demonstrate in particular that our methodology can be used to show the absence of SO(8) [and the presence of SU(3)] using the nonlinear sigma model.

II. METHODOLOGY

To detect the Lie algebra, we take advantage of the fact that the symmetry is continuous and locally defined. In the presence of a symmetry, an infinitesimal transformation of the fields of the form $\phi_i \to \phi_i' = \phi_i + \epsilon T_{ij}\phi_j$ leads to a transformation of the effective action of $\mathcal{O}(\epsilon^2)$:

$$V \to V + \epsilon\left(\frac{\partial V}{\partial \phi_i}\delta\phi_i + \frac{\partial V}{\partial(\partial_\mu\phi_i)}\delta\partial_\mu\phi_i\right) + \mathcal{O}(\epsilon^2)$$
$$\Rightarrow V \to V + \mathcal{O}(\epsilon^2). \tag{1}$$

The $\mathcal{O}(\epsilon^2)$ and higher terms remain as the Lie algebra lives in the tangent space of the Lie group's manifold. They are (for simplicity focusing on a single multiplet without derivative interactions):

$$V \to V + \epsilon^2\frac{\partial^2 V}{\partial \phi_i^{\,2}}(\delta\phi_i)^2 + \mathcal{O}(\epsilon^3). \tag{2}$$

A neural network can be used to interpolate a dataset and make predictions for the transformed fields. Then, if the symmetry is present, we should find

$$(\Delta V)_{\mathrm{NN}} \equiv \left|\frac{V_{\mathrm{NN}}(\phi_i') - V(\phi_i)}{V(\phi_i)}\right| = \epsilon^2\frac{V''_{\mathrm{NN}}(\phi_i)}{V(\phi_i)}(\delta\phi_j)^2 + \frac{E_{\%}}{100\%} + \mathcal{O}(\epsilon^3), \tag{3}$$

where

$$E_{\%} = \left|\frac{V_{\mathrm{NN}}(\phi_i) - V(\phi_i)}{V(\phi_i)}\right| \times 100\% \tag{4}$$

is the absolute percentage error of the NN on the validation set, ϕ is a datapoint in the validation set, ϕ' its image under the transformation to be tested, and $V_{\mathrm{NN}}(\phi_i')$ is the NN prediction of the transformed field. As ϕ_i is part of the dataset, $V(\phi_i)$ is known and does not need to be predicted by the network. Importantly, the ϵ^2 scaling is independent of the normalization, and $V''_{\mathrm{NN}}(\phi_i)(T_{ij}\phi_j)^2/V(\phi_i) \sim 1$ approximates the *a priori* unknown coefficients in the expansion.

III. MODELS

We use as inspiration a BSM scenario of a new scale of spontaneous symmetry breaking (SSB) that leaves behind Nambu-Goldstone boson (NGB) fields.[1] These would generically be the lightest fields and a reasonable guess as the earliest indication of the new physics. The symmetries exhibited by NGB interactions would then be a probe of the structure of the theory at or above the symmetry breaking scale.

The interactions of the NGBs are parametrized by low-energy effective theories of spontaneously broken symmetries. We will focus on two such benchmark models, the linear and nonlinear Σ models.

A. Nonlinear Σ model

The nonlinear Σ model (NLΣM) is given by:

$$\mathcal{L} = \frac{f^2}{4}\mathrm{tr}(\partial_\mu\Sigma\partial^\mu\Sigma^\dagger), \qquad \Sigma = \exp(i\pi^a t^a/f). \tag{5}$$

which has a nonlinearly realized $SU(N)_L \times SU(N)_R$ chiral symmetry and a preserved $SU(N)_F$ flavor symmetry below the SSB scale f.

As the flavor symmetry is manifest order by order in f, we can expand to $\mathcal{O}(1/f^2)$ to obtain:

$$V = -\frac{1}{24f^2}[\partial_\mu\pi^a\pi^b\partial^\mu\pi^c\pi^d - \pi^a\partial_\mu\pi^b\partial^\mu\pi^c\pi^d]\mathrm{tr}(t^a t^b t^c t^d). \tag{6}$$

The pions of (6) are in the adjoint representation of $SU(N)_F$, and transform as

$$\pi^a \to \pi^a + \epsilon f^{abc}\Theta^b\pi^c, \tag{7}$$

[1]In a realistic model, explicit breaking of the nonlinearly realized symmetry would lift the NGB masses, and these would actually be pNGBs. We leave a study of explicit breaking to future work, and will refer to these fields as NGBs. Note that the linearly realized flavor symmetry studied here could remain preserved under such explicit breaking.

where $\epsilon\Theta^a$ gives a set of infinitesimal transformation parameters. The f^{abc} are the structure constants of $SU(N)$

$$[t^a, t^b] = if^{abc}t^c, \tag{8}$$

which form the Lie algebra. Under the transformation in (7), the potential changes as $V \to V + \mathcal{O}(\epsilon^2)$.

Note that for $N \leq 2$, $SU(N) \simeq SO(N-1)$. Our goal is to identify the $SU(N)$ flavor symmetry of the NLΣM, and in general $SU(N)$ will not be isomorphic to any $SO(N)$ group. Consider the NLΣM with the lowest $SU(N > 2)$ flavor symmetry. In this case the pions form an 8-plet in the adjoint representation of $SU(3)$, but could also be rotated under $SO(8)$. Acting an $SO(8)$ transformation on these pions gives

$$\pi_i \to R_{ij}\pi_j = \pi_i + \epsilon T^a_{ij}\Theta^a\pi_j + \mathcal{O}(\epsilon^2), \tag{9}$$

where T^a are the generators of the $SO(8)$ Lie algebra. This yields

$$V \to V + \mathcal{O}(\epsilon), \tag{10}$$

as one would expect for an infinitesimal transformation not associated with a symmetry of the theory. The ability to disentangle $SU(3)$ from $SO(8)$ is thus required to detect the correct symmetry present in the NLΣM.

To summarize, the NLΣM with an $N = 3$ flavor symmetry changes as:

$$\begin{aligned} V &\xrightarrow{SU(3)} V + \mathcal{O}(\epsilon^2) \text{ symmetry present} \\ V &\xrightarrow{SO(8)} V + \mathcal{O}(\epsilon) \text{ symmetry absent} \end{aligned} \tag{11}$$

under $SU(3)$ and $SO(8)$ transformations of the π^a fields. This behavior will be exploited in our symmetry detection strategy below.

B. Linear Σ model

The same symmetry pattern $SU(N)_L \times SU(N)_R \to SU(N)_F$ can be described by the linear Σ model (LΣM), given by

$$\begin{aligned} \mathcal{L} = {} & \mathrm{tr}(\partial_\mu\Sigma\partial^\mu\Sigma^\dagger) + m_\Sigma^2\mathrm{tr}\Sigma\Sigma^\dagger + (\mu_\Sigma \det\Sigma + \text{H.c.}) \\ & - \frac{\lambda}{2}(\mathrm{tr}\Sigma\Sigma^\dagger)^2 - \frac{\kappa}{2}\mathrm{tr}\Sigma\Sigma^\dagger\Sigma\Sigma^\dagger, \end{aligned} \tag{12}$$

where

$$\Sigma_{ij} = \frac{\varphi + i\eta'}{\sqrt{2N}} + X^a t^a_{ij} + i\pi^a t^a_{ij}. \tag{13}$$

Working from the assumption that the NGB fields will be the lightest, we integrate out the heavy X, φ fields associated with unbroken generators. For simplicity, we assume sufficient symmetry breaking effects to lift the mass of the η' field enough so that it may also be integrated out.[2] This leaves only the pion field interactions:

$$\begin{aligned} \mathcal{L} = {} & \frac{1}{2}\partial_\mu\pi^a\partial^\mu\pi^a + \frac{1}{2}m_\Sigma^2\pi^a\pi^a + \frac{\lambda}{8}(\pi^a\pi^a)^2 \\ & + \frac{\kappa}{2}\mathrm{tr}(t^a t^b t^c t^d)\pi^a\pi^b\pi^c\pi^d. \end{aligned} \tag{14}$$

This potential is again invariant under the $SU(N)$ transformations of (7). Unlike the NLΣM, however, the potential in (14) is also invariant under an $SO(8)$ symmetry for $N = 3$ flavors:

$$\begin{aligned} V_{\mathrm{L\Sigma M}} &\xrightarrow{SU(3)} V_{\mathrm{L\Sigma M}} + \mathcal{O}(\epsilon^2) \\ V_{\mathrm{L\Sigma M}} &\xrightarrow{SO(8)} V_{\mathrm{L\Sigma M}} + \mathcal{O}(\epsilon^2). \end{aligned} \tag{15}$$

We therefore expect to be able to detect the presence of both symmetries. It will be useful to contrast a symmetry transformation with a nonsymmetric transformation. For this purpose we use the simple transformation:

$$\pi_i \to \pi_i + \epsilon_{ij}\pi_j, \qquad \epsilon_{ij} = \epsilon \times \frac{1}{n^2} \quad \text{for all } i, j, \tag{16}$$

where n is the number of π fields. This transformation does not correspond to any symmetry and changes the potential as $V_{\mathrm{L\Sigma M}} \to V_{\mathrm{L\Sigma M}} + \mathcal{O}(\epsilon)$. We will refer to this transformation as arb(8) in the rest of this paper.

IV. NEURAL NETWORK

The methodology proposed above uses a sequential feed-forward neural network to perform regression. As a result of the universal approximation theorem (UAT) [22,23], there is no theoretical limit to the accuracy with which a neural network with a single hidden layer and enough neurons can approximate any function. Moreover, as was recently demonstrated in [3], hidden layers can increase the interpolating abilities of the NN (the LΣM (13) and NLΣM (6) SU(3) potentials contain 80 and 143 terms from 8 and 16 dimensional input respectively). In this section we report on neural network's architecture and hyperparameters used in the analysis below. We motivate these choices in the Supplemental Material [24].

[2]If (13) describes the low energy theory behavior of QCD-like confinement of some non-Abelian gauge field, the corresponding η' would generically acquire a mass of the order of $m_{X,\varphi}$ due to explicit $U(1)_A$ breaking from instanton effects.

TABLE I. Neural network hyperparameters.

Hyperparameter	Value
Hidden layers	8
Neurons/layer	512
Optimizer	Adam
Learning rate	10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.99$
Loss function	MAPE
Training epochs	215
Training set size	$10^6 \times 0.9$
Batch size	16

To create our neural networks we used the Keras [25] library. The neural networks used had 8 hidden layers with 512 neurons, with hyperparameters as in Table I but we observed no strong dependence on this architecture. We found the best performance with an adaptive learning rate activation function with a small initial learning rate. No markers of overtraining were observed.

The training data was generated using uniform sampling in $|\phi|^{1/4}$, where $\phi = \{\pi, \partial_\mu \pi\}$ represents an input field. This distribution of input points was chosen to get an approximately Gaussian distribution in $V(\phi)$. We found that the network's $\bar{E}_{\%}$ performance scaled monotonically with the training set size, as expected. Notably, the performance was inversely correlated with batch size >8, which we attribute to the network effectively averaging out important features of the manifold.

V. SYMMETRY DETECTION

After training, we use the neural network to predict $(\Delta V)_{\mathrm{NN}}$ (3) for the validation data. In the presence of a symmetry, the converged neural network should predict $(\Delta V)_{\mathrm{NN}} \propto \epsilon^n$ with $n \geq 2$ at leading order in ϵ; in its absence, the leading term is $n = 1$. We can therefore deduce the presence of a symmetry from the absence of linear scaling for a large enough ϵ-range of predictions.

The noise due to the neural network loss function is typically correlated with the magnitude of the input vectors $|\phi|$ and depends on details of the sampling.[3] As the magnitude of the transformation ϵ is chosen independently of the input data, the neural network noise is in principle uncorrelated with ϵ. Then, the "error" in the prediction (3) for a converged network is given by

$$\begin{aligned}\text{error} &= (\Delta V)_{\mathrm{NN}} - (\Delta V)_{\mathrm{truth}} = \frac{V_{\mathrm{NN}}(\phi') - V(\phi')}{V(\phi)} \\ &\sim (1+\epsilon^n)\frac{V_{\mathrm{NN}}(\phi) - V(\phi)}{V(\phi)} + \mathcal{O}(\epsilon^{n+1}) \\ &= (1+\epsilon^n)\frac{\bar{E}_{\%}}{100\%} + \mathcal{O}(\epsilon^{n+1}) \end{aligned} \tag{17}$$

[3]This can in principle be utilized by only computing $(\Delta V)_{\mathrm{NN}}$ on data points ϕ for which the NN error $V_{\mathrm{NN}}(\phi) - V(\phi)$ is smaller than some tolerance $\delta < \bar{E}_{\%}/100\%$.

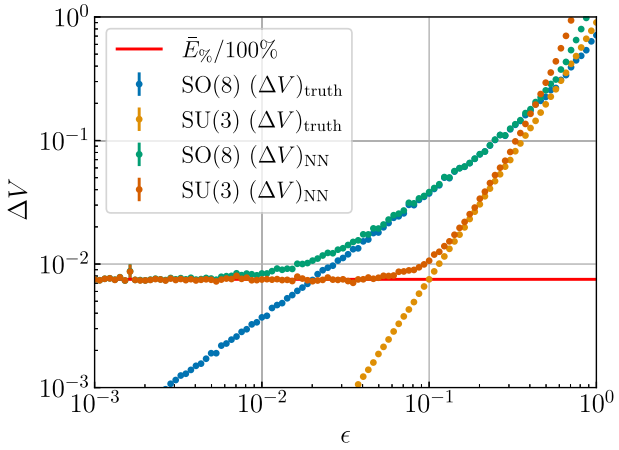

FIG. 1. The value of $(\Delta V)_{\mathrm{truth}}$ or $(\Delta V)_{\mathrm{NN}}$ as a function of ϵ for the NLΣM (6). The horizontal red line indicates the converged $\bar{E}_{\%}/100\%$, which corresponds to the expected noise floor for $(\Delta V)_{\mathrm{NN}}$. Near $\epsilon \to 1$ higher order terms in the expansion (3) become important. Error bars on the data points correspond to the standard error on the mean of ΔV obtained from all transformations performed with a given ϵ. Note that the error bars on most data points are very small.

We point out in particular that in the case of a symmetry transformation, no linear scaling is introduced in the error. Furthermore, we expect the scaling to become flat in ϵ for $\epsilon^n \lesssim \bar{E}_{\%}/100\%$.

We demonstrate the scaling of our converged network with a simple polynomial fit $(\Delta V)_{\mathrm{NN}} = \sum_{i=0}^{2} a_i \epsilon^i$ on a test set of 100 evenly spaced points in logspace in the interval $\epsilon = [10^{-3}, 1]$. For each value in ϵ, we find ΔV by averaging over 2000 transformations where the original points in field space are randomly chosen from our validation set. We plot the data points corresponding to the obtained values for $(\Delta V)_{\mathrm{NN}}$ and $(\Delta V)_{\mathrm{truth}}$ from SO(8) and SU(3) transformations in the NLΣM in Fig. 1. We give the resulting fits for both models (NLΣM and LΣM) in Table II.

Both Fig. 1 and the fit coefficients in Table II demonstrate that there is a constant error in $(\Delta V)_{\mathrm{NN}}$ which approximately corresponds to the value of $\bar{E}_{\%}$ for the network. Even by eye one can identify the linear or quadratic scaling in ΔV from Fig. 1. We find that we can correctly show that $a_1 \ll a_2$ for the SU(3) transformations in the NLΣM, and for both SU(3) and SO(8) in the LΣM model. We also correctly exclude $a_1 = 0$ for SO(8) in the NLΣM, and for the arbitrary transformation (16) in the LΣM.

We also check that linear scaling in ϵ doesn't appear for the ϵ range we consider. To test this, we construct a sliding window in ϵ with a width corresponding to an order of magnitude in logspace. On this window we evaluate our simple polynomial fit on our data points for

TABLE II. Polynomial fit $(\Delta V)_{\mathrm{NN}} = \sum_{i=0}^{2} a_i \epsilon^i$ over the full interval $\epsilon = [10^{-3}, 1]$ quoted with 1σ error bars. It is seen that $a_0 \sim \bar{E}_{\%}/100\%$ and a_1 is nonzero in the absence of symmetry.

Model	$\bar{E}_{\%}$	$\{T^a\}$	Truth?	a_0	a_1	a_2
LΣM	0.05%	SO(8)	✓	$(4.18 \pm 0.02) \times 10^{-4}$	$(0.00 \pm 9.07) \times 10^{-5}$	0.390 ± 0.001
		SU(3)	✓	$(4.14 \pm 0.02) \times 10^{-4}$	$(0.00 \pm 1.13) \times 10^{-4}$	0.588 ± 0.001
		arb(8)		$(2.19 \pm 0.02) \times 10^{-4}$	0.4001 ± 0.0001	$(6.50 \pm 0.02) \times 10^{-2}$
NLΣM	0.7%	SO(8)		$(6.74 \pm 0.03) \times 10^{-3}$	0.215 ± 0.002	0.691 ± 0.008
		SU(3)	✓	$(6.71 \pm 0.03) \times 10^{-3}$	$(0.00 \pm 1.02) \times 10^{-3}$	0.758 ± 0.005

ΔV. In Fig. 2 we plot the resulting values for a_1 as a function of ϵ. In both the NLΣM and LΣM we find that for transformations that preserve a symmetry, both $(\Delta V)_{\mathrm{truth}}$ and $(\Delta V)_{\mathrm{NN}}$ are consistent with $a_1 = 0$ for all $\epsilon \lesssim 1$, whereas for other transformations $a_1 = 0$ is excluded for a significant range of $\epsilon \ll 1$.

VI. RESULTS AND DISCUSSION

In this work we have proposed and demonstrated a method to detect a Lie group symmetry in a dataset using regression by an artificial NN. The NN was trained to replicate $V(\phi, \partial_\mu \phi)$ given training data in the form of $\{\phi, \partial_\mu \phi, V\}$. The symmetry was then tested by measuring the NN response to an $\mathcal{O}(\epsilon)$ transformation of the input fields according to the Lie algebra associated with the Lie group symmetry, effectively augmenting the dataset. We used this method to test for $SO(8)$ and $SU(3)$ symmetries in the NLΣM and LΣM, see Fig. 1. As expected, we found the NLΣM is symmetric under $SU(3)$ transformations, but is not invariant under $SO(8)$. For the LΣM, we detected the presence of both $SU(3)$ and $SO(8)$ symmetries.

The method presented here takes advantage of the fact that the Lie algebra lives in the tangent space of the group's manifold. This mitigates the importance of perfect interpolation as well as exact invariance under the full symmetry group: a symmetric system's true potential will not be exactly invariant under the Lie algebra transformation, but instead will exhibit $\mathcal{O}(\epsilon^2)$ scaling. In contrast, a system that lacks the symmetry will instead exhibit linear $\mathcal{O}(\epsilon)$ scaling. By ruling out $\mathcal{O}(\epsilon)$ scaling, we can rule out the absence of a symmetry.

The power of the neural network lies in the ability to extend this method to more realistic scenarios in which the symmetry is obscured. The next steps are to apply this technique to recover the same symmetry from more realistic data limited by minimal experimental signals or contaminated by noise. Data from more realistic experimental signals would not in general provide an ordering for the NGB fields. The generators of $SU(N)$ do not commute with the operator that shuffles these fields, and so

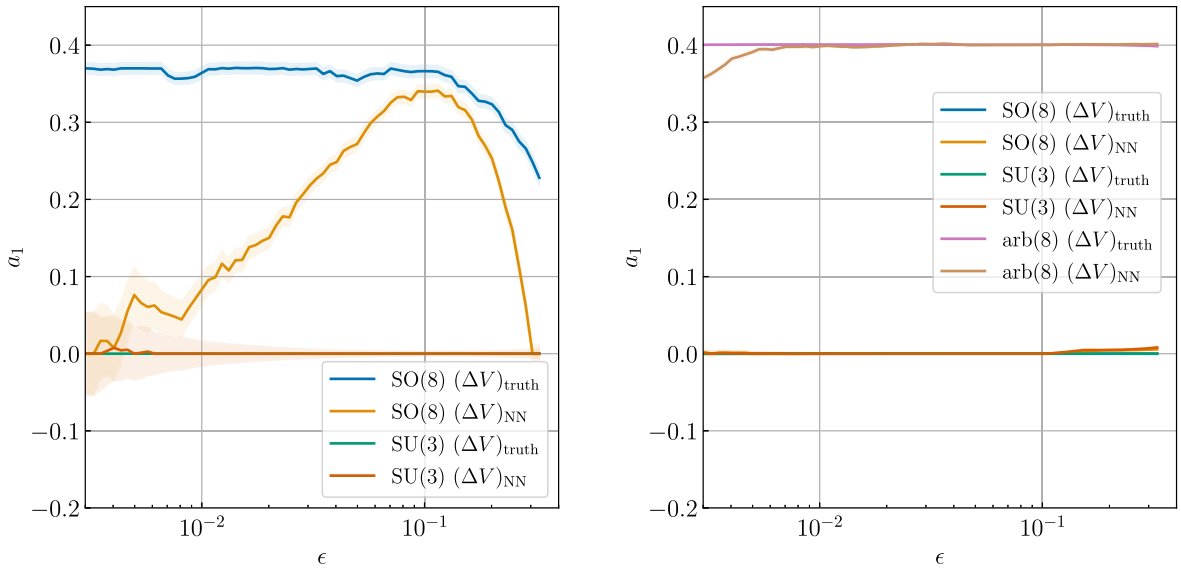

FIG. 2. Value of the coefficient a_1 from the polynomial fit $(\Delta V)_{\mathrm{NN}} = \sum_{i=0}^{2} a_i \epsilon^i$ in the NLΣM (6) (left) and LΣM (13) (right) applied on a sliding window of ϵ. The size of the window is one order of magnitude in ϵ. We plot the value of a_1 against the median value of ϵ used in the fit. The shaded region corresponds to $\pm 1\sigma$ for the fit parameter.

this method would only recover the symmetry in one of the $(N^2-1)!$ combinations of shuffled NGB fields.[4] For $SU(3)$, we are able to reorder the shuffled fields by exploiting properties of members of the Cartan subalgebra, T_3 and T_8. This trick may be formalized and extended to general $SU(N)$ or even general Lie groups, but we leave this for future study. In future work we will also study the use of this method to recover approximate symmetries in the presence of explicit breaking.

[4]We note that the possibility of indistinguishable pNGB fields is not particularly worrisome. The assumption that these fields are massive implies the presence of explicit symmetry breaking of the SSB group, generally allowing the NGB fields to be distinguished. This is the case for pions and kaons transforming under the $SU(3)$ flavor symmetry of QCD and played a key role in the discovery of the 8-fold way.

ACKNOWLEDGMENTS

The authors thank Steve Abel, Ben Allanach, Juan C. Criado, Jeff Dror, Sven Krippendorf, Dalimil Mazáč, and Michael Spannowsky for useful discussions. D. Croon thanks the Aspen Center for Physics (supported by NSF Grant No. PHY-1607611) and the Galileo Galilei Institute for hospitality during the completion of this paper. D. Cutting is supported by the Academy of Finland Grants No. 328958 and No. 345070. D. Croon and R. Houtz are supported by the STFC under Grant No. ST/T001011/1.

[1] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962).
[2] I. Low, Phys. Rev. D **91**, 105017 (2015).
[3] I. Chahrour and J. D. Wells, arXiv:2111.14788.
[4] T. J. Sejnowski, P. K. Kienker, and G. E. Hinton, Physica (Amsterdam) **22**, 260 (1986).
[5] W. Konen and C. von der Malsburg, Unsupervised symmetry detection: A network which learns from single examples, in *Proceedings of the International Conference on Artificial Neural Networks* (North-Holland, Bochum, 2008).
[6] S. Tsogkas and I. Kokkinos, *European Conference on Computer Vision* (Springer, 2012).
[7] R. Nagar and S. Raman, Pattern Recognit. **107**, 107483 (2020).
[8] J. K. George, C. Soci, M. Miscuglio, and V. J. Sorger, Sci. Rep. **11**, 1 (2021).
[9] T. Cohen and M. Welling, arXiv:1612.08498.
[10] M. van der Wilk, M. Bauer, S. T. John, and J. Hensman, Adv. Neural Inf. Process. Syst. **31**, 9938 (2018).
[11] G. Benton, M. Finzi, P. Izmailov, and A. G. Wilson, arXiv:2010.11882.
[12] N. Dehmamy, R. Walters, Y. Liu, D. Wang, and R. Yu, arXiv:2109.07103.
[13] D. W. Romero and S. Lohit, arXiv:2110.10211.
[14] N. Mourdoukoutas, M. Federici, G. M. Pantalos, M. van der Wilk, and V. Fortuin, arXiv:2107.09301.
[15] A. Zhou, T. Knowles, and C. Finn, arXiv:2007.02933.
[16] B. M. Dillon, G. Kasieczka, H. Olischlager, T. Plehn, P. Sorrenson, and L. Vogel, arXiv:2108.04253.
[17] H.-Y. Chen, Y.-H. He, S. Lal, and M. Z. Zaz, arXiv:2006.16114.
[18] H.-Y. Chen, Y.-H. He, S. Lal, and S. Majumder, Phys. Lett. B **817**, 136297 (2021).
[19] K. Desai, B. Nachman, and J. Thaler, arXiv:2112.05722.
[20] S. Krippendorf and M. Syvaeri, arXiv:2003.13679.
[21] G. Barenboim, J. Hirn, and V. Sanz, SciPost Phys. **11**, 014 (2021).
[22] G. Cybenko, Math. Control Signal Systems **2**, 303 (1989).
[23] K. Hornik, Neural Netw. **4**, 251 (1991).
[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.105.096030 for an optimisation study of the hyperparameters of our neural networks.
[25] F. Chollet *et al.*, Keras, https://keras.io (2015).