This is the accepted manuscript made available via CHORUS. The article has been published as:

Machine learning of independent conservation laws through neural deflation

Wei Zhu, Hong-Kun Zhang, and P. G. Kevrekidis

Phys. Rev. E **108**, L022301 — Published 18 August 2023

DOI: 10.1103/PhysRevE.108.L022301

Machine learning of independent conservation laws through neural deflation

Wei Zhu,[1,*] Hong-Kun Zhang,[1] and P. G. Kevrekidis[1]

[1] *Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003-4515, USA*

(Dated: August 2, 2023)

We introduce a methodology for seeking conservation laws within a Hamiltonian dynamical system, which we term "neural deflation". Inspired by deflation methods for steady states of dynamical systems, we propose to iteratively train a number of neural networks to minimize a regularized loss function accounting for the necessity of conserved quantities to be *in involution* and enforcing functional independence thereof consistently in the infinite-sample limit. The method is applied to a series of integrable and non-integrable lattice differential-difference equations. In the former, the predicted number of conservation laws extensively grows with the number of degrees of freedom, while for the latter, it generically stops at a threshold related to the number of conserved quantities in the system. This data-driven tool could prove valuable in assessing a model's conserved quantities and its potential integrability.

I. INTRODUCTION

The topic of identification of conservation laws and of the potential integrability of a Hamiltonian dynamical system has been central to both classical [1, 2], and quantum systems. In particular, it is expected for a d-dimensional dynamical system that there will generically exist some $d/2$ Poisson-commuting (i.e., *in involution*) conserved quantities to ensure integrability in the Liouville sense. Since the relevant settings arise in a wide variety of physical applications including, but not limited to, optical, atomic, material, fluid and plasma models [3–7], such features remain a central and widely studied topic.

This theme has a time-honored history and there have been numerous methods, including ones based on Painlevé property [8], as well as ones based on Lyapunov exponents (see, e.g., [9, 10] and [11] for an associated recent discussion). Nevertheless, over the past few years, there has been an extensive effort in this direction based on the premise of data-driven methods, enabling the identification of conservation laws via a variety of machine-learning techniques. Relevant methodologies have extended from employing specialized neural architectures to designing customized loss functions aimed at learning symmetries [12, 13] and conservation laws [14–19] within a given dynamical system. They also span the AI Poincaré approach learning conservation laws from trajectories [20] and discovering hidden symmetries [21] to the most recent and state-of-the-art approach of learning such conservation laws from the system's (differential) equations [22]. This wide range of efforts indicates the significance and potential of such methods, despite possible limitations. Indeed, we are not aware of methods proposed so far, able to detect the progressive increase of conservation laws, especially when the number of degrees of freedom increases. We are not familiar with efforts to detect the integrability of the associated system for a large number of degrees of freedom. Indeed, when used for integrable systems, the methods typically identify a few conservation laws [19, 22], which are argued to be relevant (e.g., physically).

Our aim in the present work is to present a method for identifying the number of conservation laws of a system, with a view to large(r) numbers of differential equations. We are motivated by the notion of *deflation* for steady states of partial differential equations [23], whereby once a stationary state has been identified, subsequent iteration steps weigh against proximity to such a state, thus discovering additional ones. Here, we devise a data-driven methodology in which the regularized loss function accounts for two central features (in our effort to seek additional conservation laws). Firstly, these must be in involution with earlier ones and furthermore, while they are not required to ensure point-wise *orthogonality* in the gradients (as in [22]), they do need to enforce *linear independence* in the gradients from earlier ones to achieve *functional independence*. Iteratively accounting for these features, we showcase that not only can we capture the appropriate number of conservation laws for systems previously benchmarked such as isotropic and anisotropic oscillators and the three-body problem. We are also able to do so for fundamental differential-difference integrable and non-integrable systems, such as the (integrable) Toda lattice [24] and (the associated non-integrable, famous) Fermi-Pasta-Ulam-Tsingou system [25] and similarly for the integrable Calogero model [26, 27], as well as the discrete sine-Gordon equation [5] (of wide physical relevance to coupled torsion pendula and superconducting Josephson junctions).

II. METHOD

Consider a d-dimensional Hamiltonian system,

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}), \quad \mathbf{f}(\mathbf{x}) = J(\mathbf{x})\nabla H(\mathbf{x}), \quad \mathbf{x} \in D \subset \mathbb{R}^d, \quad (1)$$

where $H : D \to \mathbb{R}$ is the Hamiltonian function, $J(\mathbf{x}) \in \mathbb{R}^{d\times d}$ is an anti-symmetric matrix. The Poisson bracket for two smooth functions F and G on D takes the form

$$\{F, G\}(\mathbf{x}) = \nabla F(\mathbf{x})^T J(\mathbf{x}) \nabla G(\mathbf{x}), \quad \mathbf{x} \in D. \quad (2)$$

* weizhu@umass.edu

A *conservation law* of system (1) is characterized by the vanishing of the Poisson bracket with H. More precisely, a function $I : D \to \mathbb{R}$ is a conservation law if

$$\{I, H\}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \nabla I(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in D, \tag{3}$$

where $\mathbf{f}(\mathbf{x})$ is given by Eq. (1). A collection of K conservation laws $\{I_k\}_{k\in[K]}$, where $[K] := \{1, \cdots, K\}$, is said to be *functionally independent* if their gradients $\{\nabla I_k(\mathbf{x})\}_{k\in[K]}$ are *linearly independent* as vectors in $\mathbb{R}^d$ for all $\mathbf{x} \in D$. Furthermore, they are said to be *Poisson-commuting*, or *in involution*, if their pair-wise Poisson bracket vanishes, i.e., for all $j \neq k$, $\{I_j, I_k\} = 0$.

Our goal is to use machine learning to obtain a *maximal* set of functionally independent, Poisson-commuting conservation laws $\{I_k(\mathbf{x})\}_{k=1}^{d_0}$. If $d_0 = d/2$, then the Hamiltonian system is said to be *integrable* (in the Liouville sense.) It is worth noting that the number $d_0 \leq d/2$ is generally unknown *a priori*, and the difficulty is to determine d_0 in a principled data-driven manner.

A recent attempt to achieve this was proposed by Liu et al. [22], where they considered the canonical orthogonality condition (3), with a symplectic matrix J. Specifically, they randomly sample from $D \subset \mathbb{R}^d$ a training set $\mathcal{T}$ and a validation set $\mathcal{V}$, and parameterize each conserved quantity $I_k(\mathbf{x})$ using a neural network $I_k(\mathbf{x}; \boldsymbol{\theta}_k)$, where $\boldsymbol{\theta}_k$ are the trainable network parameters. Since d_0 is generally unknown, they simultaneously train a total of d neural networks $\{I_k(\mathbf{x}; \boldsymbol{\theta}_k)\}_{k=1}^d$ to minimize the regularized loss function

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_d; \mathcal{T}) := \frac{1}{d}\sum_{k=1}^{d} \frac{1}{|\mathcal{T}|}\sum_{\mathbf{x}\in\mathcal{T}} \ell_{\text{conserv}}[\boldsymbol{\theta}_k; \mathbf{x}] \\ + \lambda \frac{2}{d(d-1)} \sum_{k\neq l} \frac{1}{|\mathcal{T}|}\sum_{\mathbf{x}\in\mathcal{T}} R[\boldsymbol{\theta}_k, \boldsymbol{\theta}_l; \mathbf{x}].\end{aligned} \tag{4}$$

The first term in (4) is the mean of the *conservation loss* for each I_k based on the condition (3), i.e.,

$$\ell_{\text{conserv}}[\boldsymbol{\theta}_k; \mathbf{x}] := \left|\widehat{\mathbf{f}}(\mathbf{x}) \cdot \widehat{\nabla I_k}(\mathbf{x}; \boldsymbol{\theta}_k)\right|^2, \tag{5}$$

where $\widehat{\mathbf{f}}(\mathbf{x})$ and $\widehat{\nabla I_k}(\mathbf{x}; \boldsymbol{\theta}_k)$ are, respectively, the l^2-normalized vectors of $\mathbf{f}(\mathbf{x})$ and $\nabla I_k(\mathbf{x}; \boldsymbol{\theta}_k)$. The second term in (4) is a regularization to encourage functional independence among $\{I_k(\cdot; \boldsymbol{\theta}_k)\}_{k=1}^d$ by enforcing point-wise *orthogonality* among the gradients $\{\nabla I_k(\cdot; \boldsymbol{\theta}_k)\}_{k=1}^d$,

$$R[\boldsymbol{\theta}_k, \boldsymbol{\theta}_l; \mathbf{x}] := \left|\widehat{\nabla I_k}(\mathbf{x}; \boldsymbol{\theta}_k) \cdot \widehat{\nabla I_l}(\mathbf{x}; \boldsymbol{\theta}_l)\right|^2. \tag{6}$$

Once $\{I_k(\mathbf{x}; \boldsymbol{\theta}_k^*)\}_{k=1}^d$ are trained, they consider the following Jacobian matrices on the validation set $\mathcal{V}$,

$$[\nabla I_1(\mathbf{x}; \boldsymbol{\theta}_1^*), \cdots, \nabla I_d(\mathbf{x}; \boldsymbol{\theta}_d^*)] \in \mathbb{R}^{d\times d}, \quad \mathbf{x} \in \mathcal{V}, \tag{7}$$

and a maximal functionally independent subset of the learned $\{I_k(\mathbf{x}; \boldsymbol{\theta}_k^*)\}_{k=1}^d$ is obtained by identifying the largest set of columns $\mathcal{K} \subset \{1, \cdots, d\}$ from the above matrices that are linearly independent for all $\mathbf{x} \in \mathcal{V}$. The cardinality $|\mathcal{K}|$ is then declared as the number d_0 of the independent conservation laws of the system.

Although the above method offers a rough estimation of the independent conservation laws $\{I_k\}_{k=1}^{d_0}$ and their total number d_0, it suffers from several limitations. Firstly, the regularization term in (4) encourages point-wise *orthogonality* among the gradients $\{\nabla I_k(\mathbf{x}; \boldsymbol{\theta}_k)\}_{k=1}^d$. However, it is important to note that functional independence of $\{I_k(\cdot; \boldsymbol{\theta}_k)\}_{k=1}^d$ merely implies *linearly independent* gradients, which are generally *not* orthogonal. Therefore, the loss function (4) is *inconsistent* in the sense that it does not vanish even if we plug into (4) the *ground-truth* set of d_0 independent conservation laws $\{I_k(\mathbf{x})\}_{k=1}^{d_0}$. As a result, one cannot declare with high confidence that the trained networks are indeed independent conservation laws simply by examining the magnitude of the loss (4) on the validation set. Secondly, Eq. (4) does not require $\{I_k(\cdot; \boldsymbol{\theta}_k)\}_{k=1}^d$ to be *in involution*. Consequently, the number of independent conservation laws detected in [22] is sometimes much larger than $d/2$, whereas there should be *at most* $d/2$ in involution.

Neural deflation method. In light of the above issues, we propose a *neural deflation method* to iteratively learn each $I_k(\cdot; \boldsymbol{\theta}_k)$ in a principled and interpretable manner, for a general Hamiltonian system (1). The benefit of our method is that the loss function for each $I_k(\cdot; \boldsymbol{\theta}_k)$ is close to zero if and only if there exist at least k independent conservation laws in involution. Therefore, the number of the conservation laws can be determined by identifying the index $k \in \{1, \cdots, d\}$ after which there is a significant jump in the validation loss.

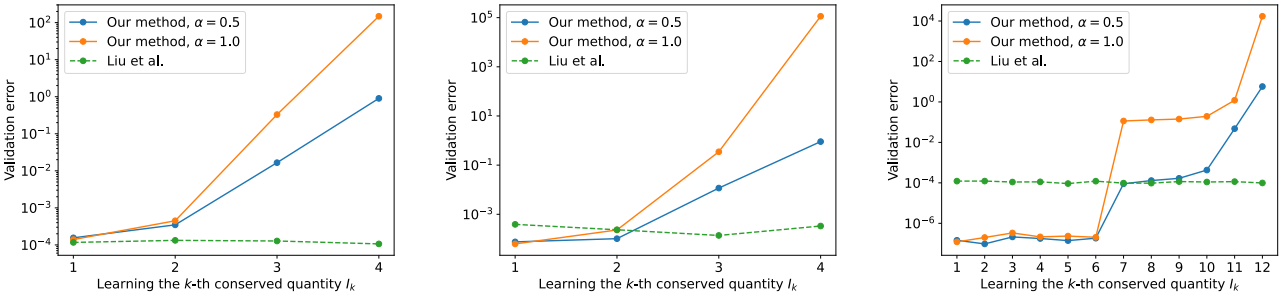
Specifically, to learn the first conservation law $I_1(\cdot; \boldsymbol{\theta}_1)$, we minimize the following loss function $\mathcal{L}_1(\boldsymbol{\theta}_1; \mathcal{T})$ based solely on the orthogonality condition (3),

$$\mathcal{L}_1(\boldsymbol{\theta}_1; \mathcal{T}) := \frac{1}{|\mathcal{T}|}\sum_{\mathbf{x}\in\mathcal{T}} \left|\widehat{\mathbf{f}}(\mathbf{x}) \cdot \widehat{\nabla I_1}(\mathbf{x}; \boldsymbol{\theta}_1)\right|^2, \tag{8}$$

In fact, since the Hamiltonian H is always a conserved quantity for (1), we can in principle directly set $I_1(\mathbf{x})$ to be $H(\mathbf{x})$ without parameterizing it as a neural network. However, training $I_1(\mathbf{x}; \boldsymbol{\theta}_1)$ based on (8) can provide us a gauge on the magnitude of the training/validation loss in order to learn and identify the subsequent conserved quantities.

We then inductively learn a sequence of conservation laws as follows. Assuming we have already obtained $K-1$ conserved quantities $\{I_k(\mathbf{x}; \boldsymbol{\theta}_k^*)\}_{k=1}^{K-1}$, where $K \geq 2$, we train the K-th conservation law $I_K(\mathbf{x}; \boldsymbol{\theta}_K)$ using the following *deflated* loss function $\mathcal{L}_K(\boldsymbol{\theta}_K; \mathcal{T})$ while fixing the learned parameters $\{\boldsymbol{\theta}_k^*\}_{k=1}^{K-1}$ of the previous networks,

$$\begin{aligned}&\mathcal{L}_K(\boldsymbol{\theta}_K; \mathcal{T}) \\ &:= \frac{1}{|\mathcal{T}|}\sum_{\mathbf{x}\in\mathcal{T}} \frac{\ell_{\text{conserv}}[\boldsymbol{\theta}_K; \mathbf{x}] + \sum_{k=1}^{K-1} \ell_{\text{inv}}[\boldsymbol{\theta}_k^*, \boldsymbol{\theta}_K; \mathbf{x}]}{K\left|\ell_{\text{ind}}\left[\boldsymbol{\theta}_K | \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{K-1}^*; \mathbf{x}\right]\right|^{\alpha}}\end{aligned} \tag{9}$$



(a) Isotropic harmonic oscillator. (b) Anisotropic harmonic oscillator. (c) 2D Three-body problem.

Figure 1: The validation losses $\{\mathcal{L}_k(\boldsymbol{\theta}_k^*;\mathcal{V})\}_{k=1}^d$ of the conserved quantities $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^d$ trained using Algorithm 1 (solid lines) for three integrable Hamiltonian systems. A significant jump in loss occurs at $k = d/2+1$ in each case, indicating accurate prediction of integrability. This trend holds across different choices of deflation strength ($\alpha = 1.0$ or $\alpha = 0.5$), with a more pronounced increase in validation loss observed for larger deflation strength. In comparison, the dashed lines show the validation losses of the conservation laws learned using Eq. (4) as proposed in [22]. These losses exhibit a similar magnitude across the d learned conservation laws, indicating the failure of the method in learning *independent and Poisson-commuting* conserved quantities.

where

$$\ell_{\text{inv}}[\boldsymbol{\theta}_k^*, \boldsymbol{\theta}_K; \mathbf{x}] := \left|\{I_k(\cdot;\boldsymbol{\theta}_k^*), I_K(\cdot;\boldsymbol{\theta}_K)\}(\mathbf{x})\right|^2$$

ensures that $I_K(\cdot;\boldsymbol{\theta}_K)$ is in involution with all previously learned $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^{K-1}$, and the loss is divided by a factor K due to having K terms in total in the numerator. The denominator is a deflation factor that enforces *functional independence* between $I_K(\cdot;\boldsymbol{\theta}_K)$ and $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^{K-1}$; more specifically,

$$\ell_{\text{ind}}[\boldsymbol{\theta}_K^* | \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{K-1}; \mathbf{x}] := \left\| \text{Proj}_{\text{span}\{\widehat{\nabla I_k}(\mathbf{x};\boldsymbol{\theta}_k^*)\}_{k\in[K-1]}^{\perp}} \widehat{\nabla I_K}(\mathbf{x};\boldsymbol{\theta}_K) \right\|^2, \quad (10)$$

where $\text{Proj}_{\text{span}\{\widehat{\nabla I_k}(\mathbf{x};\boldsymbol{\theta}_k^*)\}_{k\in[K-1]}^{\perp}} \widehat{\nabla I_K}(\mathbf{x};\boldsymbol{\theta}_K)$ denotes the projection of $\widehat{\nabla I_K}(\mathbf{x};\boldsymbol{\theta}_K)$ onto the orthogonal complement of the subspace in $\mathbb{R}^d$ spanned by $\{\widehat{\nabla I_k}(\mathbf{x};\boldsymbol{\theta}_k^*)\}_{k\in[K-1]}$. This denominator introduces a potential singularity in the loss function, penalizing the lack of independence between $\widehat{\nabla I_K}(\mathbf{x};\boldsymbol{\theta}_K)$ and $\{\widehat{\nabla I_k}(\mathbf{x};\boldsymbol{\theta}_k^*)\}_{k\in[K-1]}$. Finally, the deflation power $\alpha > 0$ serves as a hyperparameter adjusting the level of the constraint on functional independence between $I_K(\cdot;\boldsymbol{\theta}_K)$ and $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k\in[K-1]}$. See Figure 2 for a visual illustration of the learning framework.

Compared to the model (4) in [22], our model (9) has the clear advantage of being *consistent* in the infinite-sample limit. More specifically, assuming that the previously obtained $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^{K-1}$ perfectly parameterize a ground-truth set of independent conservation laws in involution and that the empirical sums $\frac{1}{|\mathcal{T}|}\sum_{\mathbf{x}\in\mathcal{T}}[\cdots]$ are replaced by the expectations $\mathbb{E}_{\mathbf{x}\sim\mu}[\cdots]$ for some probability measure μ over $D \subset \mathbb{R}^d$, then $I_K(\cdot;\boldsymbol{\theta}_K^*)$ achieves a zero loss in the infinite limit *if and only if* $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^{K}$ is a set of K independent Poisson-commuting conservation laws. Interested readers are referred to the Supplementary Material [28] for a rigorous proof of this claim.

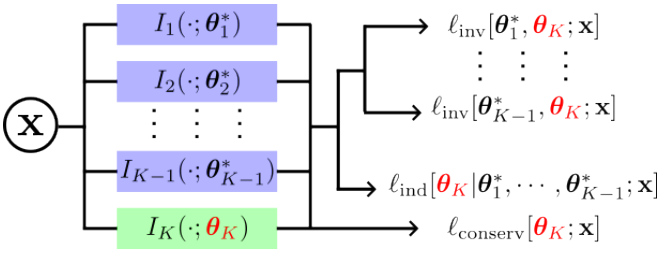
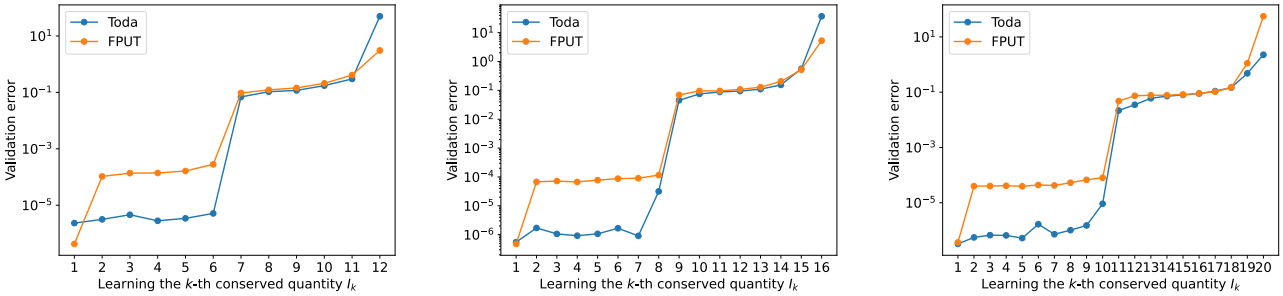


Figure 2: Visual illustration of the learning framework. Previously learned conservation laws $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k\in[K-1]}$ are fixed when training the K-th conserved quantity $I_K(\cdot;\boldsymbol{\theta}_K)$. The loss function $\mathcal{L}_K(\boldsymbol{\theta}_K;\mathcal{T})$ [Eq. (9)] is obtained by combining each loss over the training set $\mathcal{T}$.

We repeat the process until we observe a significant increase in the loss function $\mathcal{L}_K(\boldsymbol{\theta}_K^*;\mathcal{V})$ on the validation set $\mathcal{V}$, and declare, at this point, $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^{K-1}$ as a maximal set of $d_0 = K-1$ independent Poisson-commuting conservation laws of the system. Our method is summarized in Algorithm 1.

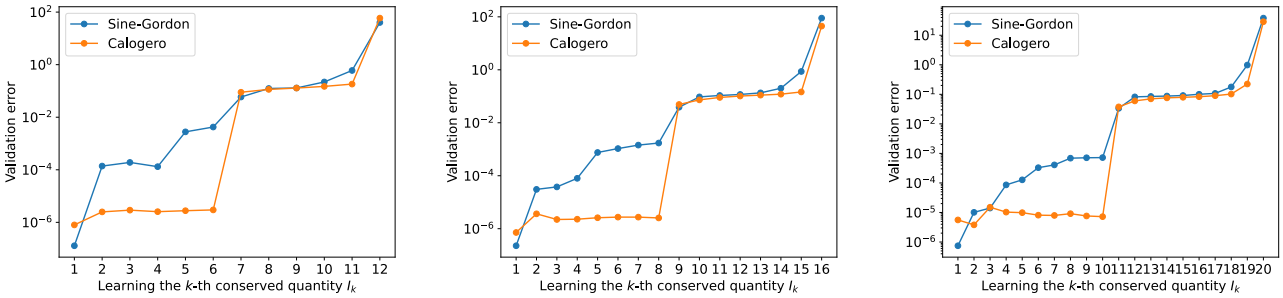
III. NUMERICAL EXPERIMENTS

We present the results of our algorithm in learning independent conservation laws of the 2D isotropic and anisotropic harmonic oscillators, the three-body problem, the Toda and the Fermi-Pasta-Ulam (FPUT) lattices, the discrete sine-Gordon system, and the Calogero's system. Mathematical and numerical details on these systems are provided in the Supplementary Material [28] (see also references [29–31] therein). The code neces-



(a) Number of lattice sites = 6. (b) Number of lattice sites = 8. (c) Number of lattice sites = 10.

Figure 3: The validation losses $\{\mathcal{L}_k(\boldsymbol{\theta}_k^*;\mathcal{V})\}_{k=1}^d$ of the conserved quantities $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^d$ trained by Algorithm 1 for the integrable Toda lattice and the non-integrable FPUT system with varying degrees of freedoms $d = 2N$, where N is the number of lattice sites. For the Toda system, the validation loss consistently exhibits a significant jump at $k = d/2+1$ for varying degrees of freedom d, whereas the jump occurs at $k = 2$ for the FPUT systems. See Section III for a detailed explanation of the results.



(a) Number of lattice sites = 6. (b) Number of lattice sites = 8. (c) Number of lattice sites = 10.

Figure 4: The validation losses $\{\mathcal{L}_k(\boldsymbol{\theta}_k^*;\mathcal{V})\}_{k=1}^d$ of the conserved quantities $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^d$ trained by Algorithm 1 for the non-integrable discrete sine-Gordon system and the integrable Calogero's system with varying degrees of freedoms $d = 2N$, where N is the number of lattice sites. For the Calogero's system, the validation loss consistently exhibits a significant jump at $k = d/2+1$ for varying degrees of freedom d, whereas the jump occurs at $k = 2$ for the sine-Gordon systems. See Section III for a detailed explanation of the results.

sary to replicate our results can be accessed at: https://github.com/mrzw731/symmetry_deflation.

The 2D isotropic/anisotropic oscillators and the three-body examples. All three systems are fully integrable in $\mathbb{R}^d$, where $d = 4$ and 12 for the harmonic oscillators and the three-body systems, respectively [22]. However, we pretend to be agnostic about their integrability, and use Algorithm 1 to obtain a maximal set of functionally independent Poisson-commuting conservation laws.

We use a 4-layer feedforward neural network with Sigmoid Linear Unit (SiLU) activations and 400 neurons per layer to parameterize each conserved quantity. To train each network, we use the ADAM optimizer [32] for 10,000 iterations with a batch size of 500. We randomly sample 200,000 phase points from the cell $[-1000, 1000]^d$ and divide them equally between the training set $\mathcal{T}$ and the validation set $\mathcal{V}$. We compare the results of setting the deflation power α in (9) to either 1.0 or 0.5.

Figure 1 displays the validation losses $\{\mathcal{L}_k(\boldsymbol{\theta}_k^*;\mathcal{V})\}_{k=1}^d$ of the learned conserved quantities $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^d$ trained using Algorithm 1. For each system, a substantial increase (of several orders of magnitude) in the validation loss occurs precisely at $k = d/2+1$, which indicates that our algorithm has accurately predicted the integrability of the systems (cf. the last line of Algorithm 1), and successfully learned a maximal set of independent conservation laws in involution. The numerical results are consistent across different choices of deflation strength, although a larger α leads to a more significant increase in validation loss at $k = d/2+1$. In comparison, we additionally present the validation losses of the conserved quantities learned using Eq. (4) as proposed in [22]. It is evident that these losses exhibit a similar magnitude across the d learned conservation laws. This observation suggests that their approach is not capable of identifying the number of *independent and Poisson-commuting* conservation laws within the system.

It is important to note that our algorithm's learned

Algorithm 1: Neural deflation method

Input: Hamiltonian system $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x})$, where $\mathbf{x} \in D \subset \mathbb{R}^d$.

Output: A maximal set $\{I_k(\cdot;\boldsymbol{\theta}_k^*)\}_{k=1}^{d_0}$ of independent conservation laws in involution.

Randomly sample a training set $\mathcal{T}$ and a validation set $\mathcal{V}$ from the phase space $D \subset \mathbb{R}^d$;
$\boldsymbol{\theta}_1^* \leftarrow \arg\min_{\boldsymbol{\theta}_1} \mathcal{L}_1(\boldsymbol{\theta}_1;\mathcal{T})$ given by Eq. (8);
$\mathcal{L}_1^{\text{val}} \leftarrow \mathcal{L}_1(\boldsymbol{\theta}_1^*;\mathcal{V})$;
$K \leftarrow 1$;
repeat
  $K \leftarrow K+1$;
  $\boldsymbol{\theta}_K^* \leftarrow \arg\min_{\boldsymbol{\theta}_K} \mathcal{L}_K(\boldsymbol{\theta}_K;\mathcal{T})$ given by (9);
  $\mathcal{L}_K^{\text{val}} \leftarrow \mathcal{L}_K(\boldsymbol{\theta}_K^*;\mathcal{V})$;
until $\mathcal{L}_K^{\text{val}}/\mathcal{L}_1^{\text{val}} >$ tol;
$d_0 \leftarrow K-1$;

conservation laws are not expected to demonstrate substantial *linear* correlations with well-established conserved quantities, such as the conservation of momentum. This is due to the fact that these laws can encompass any *nonlinear* functions derived from those quantities. Further discussion of this aspect can be found in the Supplementary Material. Nevertheless, we do note that this is an important topic for further study.

The Toda lattice and the FPUT system. We consider the integrable Toda lattice and the associated non-integrable FPUT system with different degrees of freedom $d = 2N$, where N is the number of the lattice sites ranging from 5 to 10. We use a similar experimental setup, but sample the phase points from $[-50, 50]^d$, and the network width is increased from 400 to 800. Deflation strength was only set to $\alpha = 1.0$ based on the previous experiment, and the results are shown in Figure 3. For the Toda lattice, the validation loss again significantly increases at $k = d/2+1$, (although the jump is not as "sharp", e.g., for panel (b) at $k = d/2$). This implies once again that our method accurately predicts the integrability of the Toda system and learns all the independent conservation laws. Conversely, for the FPUT system, the validation loss consistently jumps at $k = 2$ for varying d. This means our algorithm accurately predicts the non-integrability of the system and that the number of independent conservation laws remains constant across different degrees of freedom d. However, we note that the FPUT system actually has $d_0 = 2$ independent conservation laws (i.e., the momentum and H) instead of the predicted $(k-1) = 1$ learned by our algorithm. Nonetheless, the distinct behavior of the loss functions between the two systems with varying degrees of freedom highlights the potential of our algorithm in evaluating a system's integrability.

The discrete sine-Gordon system and Calogero's system. Finally, we apply our algorithm to the non-integrable discrete sine-Gordon system and the integrable Calogero's system with varying degrees of freedom d. Even though these two systems are not related, we plot the results in the same figures (Figure 4) to highlight the distinct behavior of the validation losses for an integrable vs. a non-integrable system. The numerical setup and neural architectures are exactly the same as the previous experiment. Similarly, for the (integrable) Calogero's system, the validation loss consistently exhibits a substantial increase at $k = d/2+1$ with varying degrees of freedom d. In contrast, for the (non-integrable) sine-Gordon system, the loss always jumps at $k = 2$, which is consistent with the fact that the underlying system only has *one* independent conservation law (H), regardless of the lattice size.

IV. CONCLUSIONS & FUTURE CHALLENGES

In the present work we have revisited the extensively studied in recent years topic of identifying conservation laws and, ultimately, gauging the potential integrability of a Hamiltonian model. The main contribution of the present work lies in the introduction of the technique of neural deflation. Motivated by recent developments in numerical bifurcation analysis, we propose a technique whose regularized loss function involves the involution required of the integrals of the motion and the imposition, motivated by deflation, of their linear independence. We have shown that the technique works in "standard", previously used examples such as the isotropic and anisotropic harmonic oscillator and the three-body problem [22]. Importantly, though, it successfully enables the consideration of higher-dimensional lattice nonlinear dynamical systems of both integrable (Toda, Calogero) and nonintegrable (FPUT, discrete sine-Gordon) type. In all the systems examined, we saw a distinctive increase (jump) of the loss function in the vicinity of the expected number (d_0 or just "a couple") of independent conservation laws.

Admittedly, this direction of research warrants further efforts. Examining numerous additional examples, including continuum ones, will be informative towards features such as the "sharpness" of the jump and the potential issues with capturing all the associated conservation laws (cf. the FPUT example). Another important direction is that of associating the identified quantities (and the hypersurfaces they represent) via symbolic regression to the actual conserved quantities known physically, or identified via integrability techniques in the systems of interest. Studies along this vein are currently in progress and will be reported in future work.

ACKNOWLEDGMENTS

This work is partially supported by the U.S. National Science Foundation through grants DMS-2052525 and DMS-2140982 (WZ), PHY-2110030 and DMS-2204702 (PGK), as well as DMS-2220211 and Simons Foundation -706383 (HKZ).

[1] H. Goldstein, C. P. Poole Jr., and J. L. Safko, *Classical Mechanics*, 3rd ed. (Addison-Wesley, 2001).

[2] M. Ablowitz, *Nonlinear Dispersive Waves, Asymptotic Analysis and Solitons* (Cambridge University Press, Cambridge, 2011).

[3] Y. S. Kivshar and G. P. Agrawal, *Optical Solitons: From Fibers to Photonic Crystals* (Academic Press, San Diego, California, 2003).

[4] S. Stringari and L. Pitaevskii, *Bose-Einstein Condensation* (Oxford University Press, Oxford, United Kingdom, 2003).

[5] T. Dauxois and M. Peyrard, *Physics of Solitons*, 1st ed. (Cambridge University Press, 2006).

[6] E. Infeld and G. Rowlands, *Nonlinear Waves, Solitons and Chaos* (Cambridge University Press, Cambridge, UK, 1990).

[7] M. Kono and M. Škorić, *Nonlinear Physics of Plasmas* (Springer-Verlag, Heidelberg, 2010).

[8] R. Conte, *The Painlevé property* (Springer-Verlag, New York, 1999).

[9] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory, Meccanica **15**, 9 (1980).

[10] G. Benettin, L. Galgani, A. Giorgilli, and J. Strelcyn, Lyapunov characteristic exponents for smooth dynamical systems; a method for computing all of them. Part 2: Numerical application, Meccanica **15**, 21 (1980).

[11] T. Mithun, A. Maluckov, A. Mančić, A. Khare, and P. G. Kevrekidis, How close are integrable and nonintegrable models: A parametric case study based on the salerno model, Phys. Rev. E **107**, 024202 (2023).

[12] A. Decelle, V. Martin-Mayor, and B. Seoane, Learning a local symmetry with neural networks, Physical Review E **100**, 050102 (2019).

[13] R. Bondesan and A. Lamacraft, Learning symmetries of classical integrable systems (2019).

[14] Y.-i. Mototake, Interpretable conservation law estimation by deriving the symmetries of dynamics from trained deep neural networks, Physical Review E **103**, 033303 (2021).

[15] P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis, Sympnets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems, Neural Networks **132**, 166 (2020).

[16] S. J. Wetzel, R. G. Melko, J. Scott, M. Panju, and V. Ganesh, Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks, Phys. Rev. Res. **2**, 033499 (2020).

[17] F. Alet, D. Doblar, A. Zhou, J. Tenenbaum, K. Kawaguchi, and C. Finn, Noether networks: meta-learning useful conserved quantities, in *Advances in Neural Information Processing Systems*, Vol. 34, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (Curran Associates, Inc., 2021) pp. 16384–16397.

[18] S. Ha and H. Jeong, Discovering invariants via machine learning, Physical Review Research **3**, L042035 (2021).

[19] P. Y. Lu, R. Dangovski, and M. Soljačić, Discovering conservation laws using optimal transport and manifold learning (2022).

[20] Z. Liu and M. Tegmark, Machine learning conservation laws from trajectories, Phys. Rev. Lett. **126**, 180604 (2021).

[21] Z. Liu and M. Tegmark, Machine learning hidden symmetries, Phys. Rev. Lett. **128**, 180201 (2022).

[22] Z. Liu, V. Madhavan, and M. Tegmark, Machine learning conservation laws from differential equations, Physical Review E **106**, 045307 (2022).

[23] P. E. Farrell, A. Birkisson, and S. W. Funke, Deflation techniques for finding distinct solutions of nonlinear partial differential equations, SIAM J. Sci. Comput. **37**, A2026 (2015).

[24] M. Toda, *Theory of nonlinear lattices* (Springer-Verlag, Berlin, 1981).

[25] G. Gallavotti, *The Fermi–Pasta–Ulam Problem: A Status Report* (Springer-Verlag, Berlin, Germany, 2008).

[26] F. Calogero, Solution of the one-dimensional n-body problems with quadratic and/or inversely quadratic pair potentials, Journal of Mathematical Physics **12**, 419 (1971).

[27] J. Moser, Three integrable hamiltonian systems connected with isospectral deformations, Advances in Mathematics **16**, 197 (1975).

[28] For a proof of loss consistency of our proposed model, please refer to the supplementary material available at `http://`. additionally, it also provides a comprehensive background and complete description of the dynamical systems considered herein.

[29] M. Henón, Integrals of the toda lattice, Physical Review B **9**, 1921 (1974).

[30] E. Fermi, P. Pasta, S. Ulam, and M. Tsingou, *Studies of the nonlinear problems*, Tech. Rep. (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 1955).

[31] A. D. Polyanin and V. F. Zaitsev, *Handbook of Nonlinear Partial Differential Equations: Exact Solutions, Methods, and Problems* (Chapman and Hall/CRC, 2003).

[32] D. P. Kingma, J. Ba, Y. Bengio, and Y. LeCun, Adam: A method for stochastic optimization, ICLR, San Diego (2015).