Machine learning independent conservation laws through neural deflation

Wei Zhu,1 Hong-Kun Zhang,1 and P. G. Kevrekidis1

1 Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003-4515, USA
(Dated: March 29, 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 techniques have extended from symplectic neural networks for identifying Hamiltonian dynamical systems from data [12], to devising neural transformations aimed at learning the symmetries of classical integrable systems [13], and from the use of Siamese neural networks [14], to optimal transport and diffusion maps approaches for manifold learning [15] (that may work beyond conservative systems). They also span the AI Poincaré approach learning conservation laws from trajectories [16] and discovering hidden symmetries [17] to the most recent and state-of-the-art approach of learning such conservation laws from the system’s (differential) equations [18]. 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 [15, 18], 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 [19], 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 [18]), 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 [20] and (the associated non-integrable, famous) Fermi-Pasta-Ulam-Tsingou system [21] and similarly for the integrable Calogero model [22, 23], 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 Hamiltonian system under the coordinate \( x \in D \subset \mathbb{R}^d \),

\[
\frac{dx}{dt} = f(x), \quad f(x) = J(x)\nabla H(x),
\] (1)

where \( H : D \to \mathbb{R} \) is the Hamiltonian function, \( J(x) \in \mathbb{R}^{d \times d} \) is an anti-symmetric matrix. The Poisson bracket
for two smooth functions $F$ and $G$ on $D$ takes the form
\[ \{F, G\} = \nabla F(x)^T J(x) \nabla G(x) \] (2)

A conservation law is characterized by the vanishing of the Poisson bracket with $H$. More precisely, a function $I : D \to \mathbb{R}$ is a conservation law of the autonomous Hamiltonian system, if
\[ \{I, H\} = f \cdot \nabla I(x) = 0, \quad \forall x \in D. \] (3)

A collection of $K$ conservation laws $\{I_k\}_{k \in [K]}$, where $[K] := \{1, \ldots, K\}$, is said to be functionally independent if their gradients $\{\nabla I_k(x)\}_{k \in [K]}$ are linearly independent as vectors in $\mathbb{R}^d$ for all $x \in D$. Furthermore, they are said to be Poisson-commuting, or in involution, if their pairwise 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 $\{I_k(x)\}_{k=1}^{d_0}$. 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. [18], where they consider the canonical orthogonality condition (3), with a symplectic matrix $J$. Specifically, they randomly sample from the phase space $D \subset \mathbb{R}^d$ a training set $T$ and a validation set $V$, and parameterize each conserved quantity $I_k(x)$ using a neural network $I_k(x; \theta_k)$, where $\theta_k$ are the trainable network parameters. Since $d_0$ is generally unknown, they simultaneously train a total of $d$ neural networks $\{I_k(x; \theta_k)\}_{k=1}^{d_0}$ to minimize the following regularized loss function
\[
\mathcal{L}(\theta_1, \ldots, \theta_d; T) := \frac{1}{d} \sum_{k=1}^{d} \ell_{\text{conserv}}[I_k(\cdot; \theta_k); T] + \lambda \frac{2}{d(d-1)} \sum_{k \neq l} R(\theta_k, \theta_l; T). \tag{4}
\]

The first term in (4) is the mean of the conservation losses for each $I_k$ based on the condition (3), i.e.,
\[
\ell_{\text{conserv}}[I_k(\cdot; \theta_k); T] := \frac{1}{|T|} \sum_{x \in T} \left| \tilde{f}(x) \cdot \nabla I_k(x; \theta_k) \right|^2, \tag{5}
\]
where $\tilde{f}(x)$ and $\nabla I_k(x; \theta_k)$ are, respectively, the $l^2$-normalized vectors of $f(x)$ and $\nabla I_k(x; \theta_k)$. The second term in (4) is a regularization to encourage functional independence among $\{I_k(\cdot; \theta_k)\}_{k=1}^{d_0}$ by enforcing point-wise orthogonality among the gradients $\{\nabla I_k(\cdot; \theta_k)\}_{k=1}^{d_0}$,
\[
R(\theta_k, \theta_l; T) := \frac{1}{|T|} \sum_{x \in T} \left| \nabla I_k(x; \theta_l) \cdot \nabla I_l(x; \theta_l) \right|^2. \tag{6}
\]

Once $\{I_k(x; \theta_k^*)\}_{k=1}^{d_0}$ are trained, they consider the following Jacobian matrices on the validation set $V$,
\[
[\nabla I_1(x; \theta_1^*), \ldots, \nabla I_d(x; \theta_d^*)] \in \mathbb{R}^{d \times d}, \quad x \in V, \tag{7}
\]
and a maximal functionally independent subset of the learned $\{I_k(x; \theta_k^*)\}_{k=1}^{d_0}$ is obtained by identifying the largest set of columns $K \subset \{1, \ldots, d\}$ of the above matrices that are consistently linearly independent for all $x \in V$. The cardinality $|K|$ is then declared as the number $d_0$ of the independent conservation laws of 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 $\{\nabla I_k(x; \theta_k)\}_{k=1}^{d_0}$. However, it is important to note that functional independence of $\{I_k(\cdot; \theta_k)\}_{k=1}^{d_0}$ 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 $\{I_k(x)\}_{k=0}^{d_0}$. 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 $\{I_k(\cdot; \theta_k)\}_{k=1}^{d_0}$ to be in involution. Consequently, the number of independent conservation laws detected in [18] 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 the neural deflation method to iteratively learn each $I_k(\cdot; \theta_k)$ in a principled and interpretable manner, for general Hamiltonian system (1). The benefit of our method is that the loss function for each $I_k(\cdot; \theta_k)$ is close to zero if 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 $k \in \{1, \ldots, d\}$ after which there is a significant jump in the validation loss.

Specifically, to learn the first conservation law $I_1(\cdot; \theta_1)$, we minimize the following loss function $\mathcal{L}_1(\theta_1; T)$ based solely on the orthogonality condition (3),
\[
\mathcal{L}_1(\theta_1; T) := \ell_{\text{conserv}}[I_1(\cdot; \theta_1); T], \tag{8}
\]
where $\ell_{\text{conserv}}$ is defined in (5), for general anti-symmetric matrix $J$. In fact, since the Hamiltonian $H$ is always a conserved quantity for (1), we can in principle directly set $I_1(x)$ to be $H(x)$ without parameterizing it as a neural network. However, training $I_1(x; \theta_1)$ 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 $\{I_k(x; \theta_k^*)\}_{k=1}^{K-1}$, where $K \geq 2$, we train the $K$-th conservation law $I_K(x; \theta_K)$ using the following deflated loss function $\mathcal{L}_K(\theta_K; T)$ while fixing the learned parameters $\{\theta_k^*\}_{k=1}^{K-1}$ of the previous networks,
\[
\mathcal{L}_K(\theta_K; T). \tag{9}
\]
Anisotropic harmonic oscillator.

Figure 1: The validation losses \( \{ \mathcal{L}_k(\theta^*_k; \mathcal{V}) \}_{k=1}^d \) of the conserved quantities \( \{ I_k(\cdot; \theta^*_k) \}_{k=1}^d \) trained by Algorithm 1 for three integrable Hamiltonian systems, where \( d \) is the number of degrees of freedom. A significant jump in loss (of several orders of magnitude) occurs at \( k = d/2 + 1 \) in each case. This jump indicates that our method has accurately predicted the integrability of the systems. This result is consistent across different choices of deflation strength, namely \( \alpha = 1.0 \) or \( \alpha = 0.5 \), although the increase in validation loss is more pronounced with larger deflation strength.

\[
\ell_{\text{conserv}}[I_K(\cdot; \theta_K); \mathcal{T}] + \sum_{k=1}^{K-1} \ell_{\text{inv}}[\theta^*_k, \theta_K; \mathcal{T}] / K
\]

where

\[
\ell_{\text{inv}}[\theta^*_k, \theta_K; \mathcal{T}] = \frac{1}{|\mathcal{T}|} \sum_{x \in \mathcal{T}} |\{ I_k(\cdot; \theta^*_k), I_2(\cdot; \theta_K) \}(x)|^2
\]

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

\[
\text{Proj}_{\text{span}\{ \nabla I_k(\cdot; \theta^*_k) \}_{k \in \{K-1\}}} \nabla I_K(\cdot; \theta_K)
\]

denotes the projection of the vector \( \nabla I_K(\cdot; \theta_K) \) onto the orthogonal complement of the subspace in \( \mathbb{R}^d \) spanned by \( \{ \nabla I_k(\cdot; \theta^*_k) \}_{k \in \{K-1\}} \). We note that for (10) to be nonzero, in contrast to Eq. (6), \( \nabla I_K(\cdot; \theta_K) \) is only required to be linearly independent with the previously learned \( \{ \nabla I_k(\cdot; \theta^*_k) \}_{k \in \{K-1\}} \) instead of being orthogonal. Finally, the deflation power \( \alpha > 0 \) is a hyperparameter adjusting the strength of the constraint on functional independence between \( I_K(\cdot; \theta_K) \) and \( \{ I_k(\cdot; \theta^*_k) \}_{k \in \{K-1\}} \).

Compared to the model (4) in [18], our model (9) has the clear advantage of being consistent in the infinite-sample limit. More specifically, assuming that the previously obtained \( \{ I_k(\cdot; \theta^*_k) \}_{k=1}^{K-1} \) perfectly parameterize a ground-truth set of independent conservation laws in involution and that the empirical sums \( \frac{1}{|\mathcal{T}|} \sum_{x \in \mathcal{T}} \cdots \) are replaced by the expectations \( E_{x \sim \mu}[\cdots] \) for some absolutely continuous probability measure \( \mu \) over the phase space \( D \subset \mathbb{R}^d \), then \( I_K(\cdot; \theta_K) \) achieves a zero loss in the infinite limit if and only if \( \{ I_k(\cdot; \theta_k) \}_{k=1}^K \) is a set of \( K \) independent Poisson-commuting conservation laws.

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

**Algorithm 1: Neural deflation method**

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

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

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

**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 problem.

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

Figure 2: The validation losses $\{L_k(\theta^*_k; V)\}_{k=1}^d$ of the conserved quantities $\{I_k(\cdot; \theta^*_k)\}_{k=1}^d$ 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 3: The validation losses $\{L_k(\theta^*_k; V)\}_{k=1}^d$ of the conserved quantities $\{I_k(\cdot; \theta^*_k)\}_{k=1}^d$ 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.

tors and the three-body systems, respectively [18]. 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 [24] 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 $T$ and the validation set $V$. We compare the results of setting the deflation power $\alpha$ in (9) to either 1.0 or 0.5.

Figure 1 displays the validation losses $\{L_k(\theta^*_k; V)\}_{k=1}^d$ of the learned conserved quantities $\{I_k(\cdot; \theta^*_k)\}_{k=1}^d$ 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, namely $\alpha = 1.0$ or $\alpha = 0.5$, although a larger $\alpha$ leads to a more significant increase in validation loss at $k = d/2 + 1$.

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 and periodic boundary conditions. We use a similar experimental setup, but sample the phase points from $[-50, 50]^d$. Deflation strength was only set to $\alpha = 1.0$ based on the previous experiment, and the results are shown in Figure 2. 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 problem. Finally, we apply our algorithm to the non-integrable discrete sine-Gordon system and the integrable Calogero’s problem with varying degrees of freedom $d$. Even though these two systems are not related, we plot the results in the same figures (Figure 3) to highlight the distinct behavior of the validation losses for an integrable vs. a non-integrable system. Similar to the previous experiment, for the (integrable) Calogero’s problem, 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 ($\hat{H}$), 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 [18]. 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.


[22] 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).
