J. Mech. Phys. Solids 170 (2023) 105127

Contents lists available at ScienceDirect

Journal of the Mechanics and Physics of Solids

journal homepage: www.elsevier.com/locate/jmps

Extracting conservative equations from nonconservative state data

Zhanchao Huang [a], Shaohan Huang [b], Junyin Li [a], Yong Wang [a,*], Hanqing Jiang [c,*]

[a] Department of Engineering Mechanics, Zhejiang University, Hangzhou, Zhejiang 310027, China
[b] Department of Endocrinology, The Affiliated Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou 310027, China
[c] Research Center for Industries of the Future and School of Engineering, Westlake University, Hangzhou 310030, China

ARTICLE INFO

Keywords:
Theory-guided data science
Conservative equation
Lagrangian
Nonconservative system
State data
Orthogonalization

ABSTRACT

Data-driven equation identification methods enable automated discovery of systems governing equations in differential, integral, or variational forms, provided with state data for autonomous systems while state and excitation data for forced systems. These methods for equation identification can undoubtedly outperform those for algebraic relation identification, however, they essentially belong to the paradigm of data fitting, viz., compulsorily establishing mathematical relations within data. In this work, we tackle a different problem of extracting hidden conservative equations only from nonconservative state data collected from randomly/deterministically excited, dissipative dynamical systems; that is, we attempt to distill intrinsic structures without any excitation information. A 3E framework is implemented, namely, embedding Euler-Lagrangian equations in systems, eliminating the influence of nonconservative factors by orthogonalization, while extracting Lagrangians (or Lagrangian densities) for discrete (or continuous) systems. Three illustrative examples, including the Duffing oscillator, the cart-pendulum system, and the Euler-Bernoulli beam, are investigated to show how this method can achieve simplicity in complexity.

1. Introduction

Recent achievements in data sciences have promoted the automation progress of scientific discovery (Weinan et al., 2021). For example, some data-driven identification methods can now automatically discover governing equations in differential, integral, or variational forms directly from simulation or experimental data (Brunton et al., 2015; Huang et al., 2020; Kaiser et al., 2018; Reinbold et al., 2020; Rudy et al., 2017; Schaeffer and McCalla, 2017; Schmidt and Lipson, 2009). Conventional data fitting methods obtain algebraic relations within data. In contrast, equation identification methods discover dynamic relations between state data and their various-order derivatives. This conversion process in system identification, from algebraic relations to dynamic relations, represents an important paradigm shift in scientific discovery (Newton et al., 1846).

Equation identification can be classified into explicit and implicit methods. Both are parallelly developed, always mutually promoted, and mutually inspired. The seminal work by Schmidt and Lipson (2009) and Brunton et al. (2015) offered an important breakthrough in the explicit branch; similar advances were made in the implicit branch by neural ordinary differential equations (Chen et al., 2018). Schmidt and Lipson (2009) leveraged genetic programming to discover symbolic relations within data. This method, however, was constrained by computational inefficiency, the curse of dimensionality, and a proneness to overfitting. Following closely, Brunton et al. (2015) introduced the assumption of parsimony, and successfully discovered differential equations

* Corresponding authors.
E-mail addresses: ypwang@zju.edu.cn (Y. Wang), hanqing.jiang@westlake.edu.cn (H. Jiang).

https://doi.org/10.1016/j.jmps.2022.105127
Received 28 September 2022; Received in revised form 3 November 2022; Accepted 3 November 2022
Available online 5 November 2022

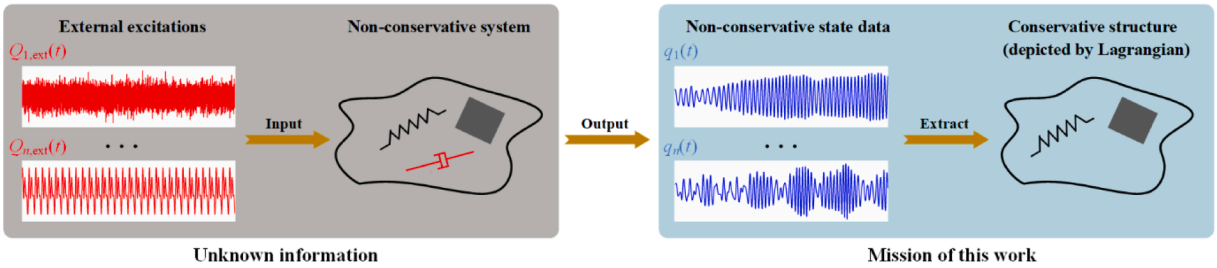

Fig. 1. Illustration of what we do not know (external excitations and nonconservative system), what we know (nonconservative state data), and what we want to know (conservative structure), presenting the unknown information and mission of this work.

using sparsity promoting optimization. This method, named SINDy, could produce interpretable models and at the same time scaled well to high-dimensional problems. However, the numerical differentiations in SINDy made it sensitive to data noise, especially when high-order derivatives were involved. To resolve this sensitivity, new methods in integral identification (Schaeffer and McCalla, 2017), weak-form identification (Reinbold et al., 2020), and variational identification (Huang et al., 2020) were developed. The first two of these methods was within the mathematical framework and could transform differential equations to integral/weak-form equations. The last method, i.e., variational identification, on the other hand, was within the framework of theory-guided data science, that is, it started from physical variational equations, which are considered the most compact description of general physical systems, as opposed to mathematical differential equations. A range of other methods, including statistical techniques such as Bayesian inference (Hirsh et al., 2022; Zhang and Lin, 2018) and bootstrap aggregating (Fasel et al., 2021) have also been used to alleviate sensitivity to data noise.

Implicit identification flourishes along the thread of neural ordinary differential equations, especially after the pioneering extensions made by Greydanus et al. (2019) and Cranmer et al. (2020). Hamiltonian neural networks (HNN) (Greydanus et al., 2019) and Lagrangian neural networks (LNN) (Cranmer et al., 2020) impose physical rigid constraints to identify differential equations of conservative systems. This is accomplished by implicitly constructing scalar functions of system states, i.e., Hamiltonian and Lagrangian, from state data. To determine generalized momenta of Hamiltonian systems (this is an essential drawback of HNN, compared to LNN), Bertalan et al. (2019) and Choudhary et al. (2021) adopted dual neural networks, in which one achieves generalized momenta while the other approximates the Hamiltonian function, with the two networks being trained simultaneously. Finzi et al. (2020), on the other hand, abandoned generalized coordinates and started directly from physical coordinates, while introducing kinematics constraints into differential equations by Lagrangian multipliers. In the aspect of nonconservative system identification, Zhong et al. (2020, Zhong et al., 2019) embedded the port-Hamiltonian formulism into neural networks, while inferring mass matrix, potential energy, dissipation matrix, and input matrix simultaneously from the given state data and pre-specified control data.

Data-driven equation identification has significantly advanced and transformed the modeling, simulation, and understanding of complex systems in scientific and industrial communities (Chen et al., 2021). It has led to extensive applications, including the tremendous potential currently in fields that lack reliable mathematical models, such as atmospheric science (Chattopadhyay et al., 2020), recognition science (Khaled et al., 2022), and epidemiology (Horrocks and Bauch, 2020), to name a few. However, despite their powerful efficacy, data-driven identification methods must be fed state data and excitation data (if there is any excitation). Thus, these methods, essentially, still belong to the paradigm of data fitting, i.e., compulsorily establishing mathematical relations within data provided. In other words, they operate within state data themselves for autonomous systems or within state and excitation data for nonconservative systems.

Here, we propose a bold question: in the absence of excitation data, is it possible to extract the associated conservative equations of randomly/deterministically excited, dissipative dynamical systems only from nonconservative state data? The fundamental concept, as shown in Fig. 1, is about seeing Essence through Surface, or eliminating the superficial fiction (excitations and dissipative components) while retaining the hidden truth (conservative equations). Note that this proposition is of important scientific and practical value. On the one hand, conservative information (Lagrangian and Hamiltonian, for instance) plays the pivotal role in physical science; we could even say that one of the primary objectives of physical science is just to find intrinsically conservative structures. On the other hand, for systems with complex or unknown excitations, it is always rather difficult or almost impossible to capture complete excitation data, such as the spatially distributed time-varying random wind load on architectural structures or unknown stimulations on micro-biological tissues. The two aspects mentioned above, thus, constitute the motivation for this work.

As preparation for this work was underway, two papers (Desai et al., 2021; Liu et al., 2021) independently explored similar subjects from the implicit point of view. Liu et al. (2021) proposed dual networks (i.e., a Lagrangian neural network plus a universal approximator network) to detect new physics from trajectories induced by unknown excitations. Similarly, Desai et al. (2021) adopted triple dual networks to simultaneously recover the underlying Hamiltonian structure as well as time-dependent forces and dissipative coefficients solely from state data. While these two works offer significant contributions, they impose rigorous limitations on excitations and dissipative components, or do not demonstrate applicability for cases with random excitations. Compared to these works, herein we aim to extract only the conservative equations, as opposed to both conservative and nonconservative components. We extract conservative equations from the explicit viewpoint rather than the implicit, with an understanding that symbolic mathematical representation is the gold standard. The state data in our approach may be captured from dynamical systems with complex excitations

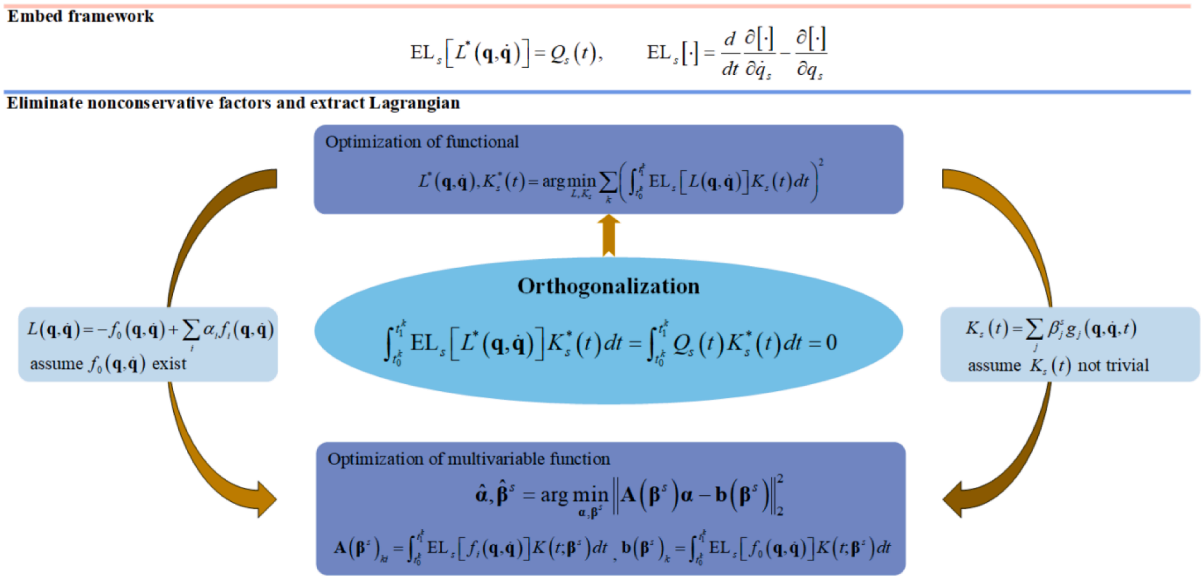

Fig. 2. The fundamental idea of the proposed method, briefly describing by three verbs, namely, embedding, eliminating, and extracting (3E). Embedding Euler-Lagrangian equations, eliminating the influences of nonconservative factors, and extracting Lagrangian.

(combination of random and deterministic excitations for instance) as well as complex dissipative components (nonlinear or even non-smooth damping).

The remainder of this work is arranged as follows: Section 2 describes the fundamental idea and the detailed methodology. In Section 3, three typical and illustrative examples, i.e., the Duffing oscillator, the cart-pendulum system, and the Euler-Bernoulli beam are investigated to demonstrate the effectiveness and accuracy of the proposed identification method, as well as its range of application with respect to nonconservative components. Section 4 presents conclusions and directions for future ideas.

2. Methodology

This work is developed within the paradigm of theory-guided data science. The fundamental idea can be condensed in a concise statement:

> ***Embed*** *Euler-Lagrangian equations in randomly/deterministically excited, dissipative systems;* ***eliminate*** *the influence of nonconservative factors by orthogonalization operations and simultaneously* ***extract*** *conservative structures by assumptions and regressions.*

The embedded structure serves as a "filter": nonconservative state data pass through the filter, and then only the effective information is retained. The understanding is that Euler-Lagrangian equations are the most fundamental conservative structures hidden in general physical systems. Thus, the embedment stated here is quite general, and almost does not limit its range of application. An orthogonalization operation lies at the heart of this method, while a regression operation gives the explicit description of the conservative structure, which is Lagrangian. As shown in Fig. 2, we embrace three words: embedding, eliminating, and extracting (3E) to demonstrate the details of this approach.

We start from state data collected from a general dynamical system. The dynamical system may be dissipative and subjected to arbitrary excitations. First, ***embed*** Euler-Lagrangian equations by which the exact relation between states and excitations is described as

$$\mathrm{EL}_s[L^*(\mathbf{q},\dot{\mathbf{q}})]=Q_s(t),s=1,\ldots,d \tag{1}$$

in which $\mathbf{q}=\{q_1,\ldots,q_d\}$ and $\dot{\mathbf{q}}=\{\dot{q}_1,\ldots,\dot{q}_d\}$ denote generalized coordinates and generalized velocities, respectively; $L^*(\mathbf{q},\dot{\mathbf{q}})$ is the true Lagrangian of the system; $\mathrm{EL}_s[\cdot]=\frac{d}{dt}\frac{\partial[\cdot]}{\partial\dot{q}_s}-\frac{\partial[\cdot]}{\partial q_s}$ is the Euler-Lagrangian operator with respect to the generalized coordinate q_s; $Q_s(t)$ is the nonconservative generalized force associated with q_s, and may be induced by excitations and dissipative components. Our goal is to identify the explicit expression of the true Lagrangian, solely from state data $(\mathbf{q},\dot{\mathbf{q}})$ and without the knowledge of $Q_s(t)$.

The stumbling block here though is that the information of nonconservative forces $Q_s(t)$ is unavailable. Different from what Liu et al. (2021) and Desai et al. (2021) adopted by penalizing the nonconservative components $Q_s(t)$ in loss functions, and simultaneously learning the neural network representation of conservative and nonconservative components, here, we try a new approach: we ***eliminate*** nonconservative components using an *orthogonalization operation*. Suppose there exists a nontrivial function (not identically equal to zero) $K_s^*(t)$ by the name of kernel function that is orthogonal to the nonconservative component $Q_s(t)$ in given time intervals $[t_0^k,t_1^k]$ $(k=1,\ldots,M)$, i.e., $\int_{t_0^k}^{t_1^k}Q_s(t)K_s^*(t)dt=0$. We multiply both sides of Eq. (1) by $K_s^*(t)$ and integrate over the given time intervals,

$$\int_{t_0^k}^{t_1^k} \mathrm{EL}_s[L^*(\mathbf{q},\dot{\mathbf{q}})]K_s^*(t)dt = \int_{t_0^k}^{t_1^k} Q_s(t)K_s^*(t)dt = 0 \tag{2}$$

Now the influence of nonconservative components is tactfully removed by the operation of weighted integration.

Hence, the true Lagrangian $L^*(\mathbf{q},\dot{\mathbf{q}})$ and the kernel function $K_s^*(t)$ can be ***extracted*** by solving the following optimization problem:

$$L^*(\mathbf{q},\dot{\mathbf{q}}), K_s^*(t) = \underset{L,K_s}{\mathrm{argmin}} \sum_{k=1}^{M} \left(\int_{t_0^k}^{t_1^k} \mathrm{EL}_s[L(\mathbf{q},\dot{\mathbf{q}})]K_s(t)dt \right)^2 \tag{3}$$

Here $L(\mathbf{q},\dot{\mathbf{q}})$ is the approximated Lagrangian, and $K_s(t)$ the corresponding kernel function. Based on the outlined 3E (embedding, eliminating, extracting) framework, we propose a data-driven method to fulfill our goal. The detailed operating steps of this method are described next.

Step 1: Collect discrete state data. State data are collected at discrete time instants within a finite time interval $[0, t_s]$, either from numerical simulations or experimental measurements.

Step 2: Construct Lagrangian and kernel functions. A Lagrangian is constructed using a linear combination of pre-specified library of functions,

$$L(\mathbf{q},\dot{\mathbf{q}};\boldsymbol{\alpha}) = -f_0(\mathbf{q},\dot{\mathbf{q}}) + \sum_{i=1}^{m} \alpha_i f_i(\mathbf{q},\dot{\mathbf{q}}) \tag{4}$$

in which, α_i, $i = 1,\ldots,m$ are to-be-determined coefficients, and $f_i(\mathbf{q},\dot{\mathbf{q}})$ are given basis functions of system states and represent candidate terms for the true Lagrangian. It is acceptable that the term $f_0(\mathbf{q},\dot{\mathbf{q}})$ exists in the true Lagrangian and its coefficient can be assumed by any number (e.g., -1 in here); such a term could be the square of generalized velocities, reflecting kinetic energy of the system. Simultaneously, we construct the kernel function for each nonconservative generalized force,

$$K_s(t;\boldsymbol{\beta}^s) = \sum_{j=1}^{n} \beta_j^s g_j^s(\mathbf{q},\dot{\mathbf{q}},t) \tag{5}$$

in which, $\beta_j^s, j = 1,\ldots,n$ are undetermined coefficients, and $g_j^s(\mathbf{q},\dot{\mathbf{q}},t)$ are given basis functions of system states and time. The selection of basis functions should be guided by domain knowledge. For instance, for systems with random excitations, $g_j^s(\mathbf{q},\dot{\mathbf{q}},t)$ can be constituted as polynomials of states due to the known orthogonality between random processes and their derivative processes (Lin and Cai, 1995; Stratonovitch, 1963). Further, we normalize each basis function $g_j^s(\mathbf{q},\dot{\mathbf{q}},t)$ to arrive at a standard deviation 1, such that the associated magnitudes β_j^s reflect their relative importance in kernel functions.

Step 3: Identify Lagrangian and kernel functions. Substitute the construction of Lagrangian in Eq. (4) and that of kernel function in Eq. (5) into the optimization problem Eq. (3), and rewrite it in matrix form,

$$\widehat{\alpha}, \widehat{\beta}^s = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}^s}{\mathrm{argmin}} \| \mathbf{A}(\boldsymbol{\beta}^s)\boldsymbol{\alpha} - \mathbf{b}(\boldsymbol{\beta}^s) \|_2^2 \tag{6}$$

in which $\mathbf{A}(\boldsymbol{\beta}^s)$ is an $M \times m$ matrix with the element in the kth-row (i.e., M time intervals) and the ith-column being $\int_{t_0^k}^{t_1^k} \mathrm{EL}_s[f_i]K_s(t;\boldsymbol{\beta}^s)dt$, and $\mathbf{b}(\boldsymbol{\beta}^s)$ is an $M \times 1$ vector with the element in the kth-row being $\int_{t_0^k}^{t_1^k} \mathrm{EL}_s[f_0]K_s(t;\boldsymbol{\beta}^s)dt$. We normalize the matrix $\mathbf{A}(\boldsymbol{\beta}^s)$ and the vector $\mathbf{b}(\boldsymbol{\beta}^s)$ by the standard deviation of $\mathbf{b}(\boldsymbol{\beta}^s)$ to avoid a trivial solution (i.e., $\boldsymbol{\beta}^s$ being zeroes). Note that in Eq. (6), if we fix one set of parameters $\boldsymbol{\beta}^s$, then the other set of parameters $\boldsymbol{\alpha}$ can be solved readily using linear algebra, i.e., $\boldsymbol{\alpha} = [\mathbf{A}^T(\boldsymbol{\beta}^s)\mathbf{A}(\boldsymbol{\beta}^s)]^{-1}\mathbf{A}^T(\boldsymbol{\beta}^s)\mathbf{b}(\boldsymbol{\beta}^s)$. Hence the optimization problem of two sets of parameters in Eq. (6) can be converted to an optimization problem of one set of parameters,

$$\widehat{\beta}^s = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}^s}{\mathrm{argmin}} \| (\mathbf{P}(\boldsymbol{\beta}^s) - \mathbf{I})\mathbf{b}(\boldsymbol{\beta}^s) \|_2^2 \tag{7}$$

in which $\mathbf{P}(\boldsymbol{\beta}^s) = \mathbf{A}(\boldsymbol{\beta}^s)[\mathbf{A}^T(\boldsymbol{\beta}^s)\mathbf{A}(\boldsymbol{\beta}^s)]^{-1}\mathbf{A}^T(\boldsymbol{\beta}^s)$ is the orthogonal projection matrix, $\mathbf{I}$ is the identity matrix. The optimization problem in Eq. (7) entails a geometric interpretation, i.e., minimizing the vertical distance from the vector $\mathbf{b}(\boldsymbol{\beta}^s)$ to the column space of the matrix $\mathbf{A}(\boldsymbol{\beta}^s)$. If the vertical distance is zero, then the kernel function is orthogonal to the nonconservative force in the given time intervals. Note that the time intervals are pre-specified in this paper for ease of presentation, but we can also determine the optimal intervals by adding one optimization nest.

By substituting the optimal solution $\widehat{\beta}^s$ into Eq. (6), the corresponding optimal solution $\widehat{\alpha}$ can be readily solved. Here, we leverage the fact that for most dynamical systems, the explicit expression of the Lagrangian is sparse in a high dimensional nonlinear function space. To obtain a parsimonious expression for the Lagrangian, we penalize the model complexity of the Lagrangian by adding the L_1-norm of $\boldsymbol{\alpha}$ to the objective function in Eq. (6), resulting in the following regularized least-squares problem

$$\widehat{\alpha} = \underset{\boldsymbol{\alpha}}{\mathrm{argmin}} \left(\| \mathbf{A}(\widehat{\beta}^s)\boldsymbol{\alpha} - \mathbf{b}(\widehat{\beta}^s) \|_2^2 + \lambda \| \boldsymbol{\alpha} \|_1 \right) \tag{8}$$

in which, $\| \cdot \|_1$ denotes the L_1-norm (James et al., 2013), and λ is a regularization parameter that balances the model complexity and

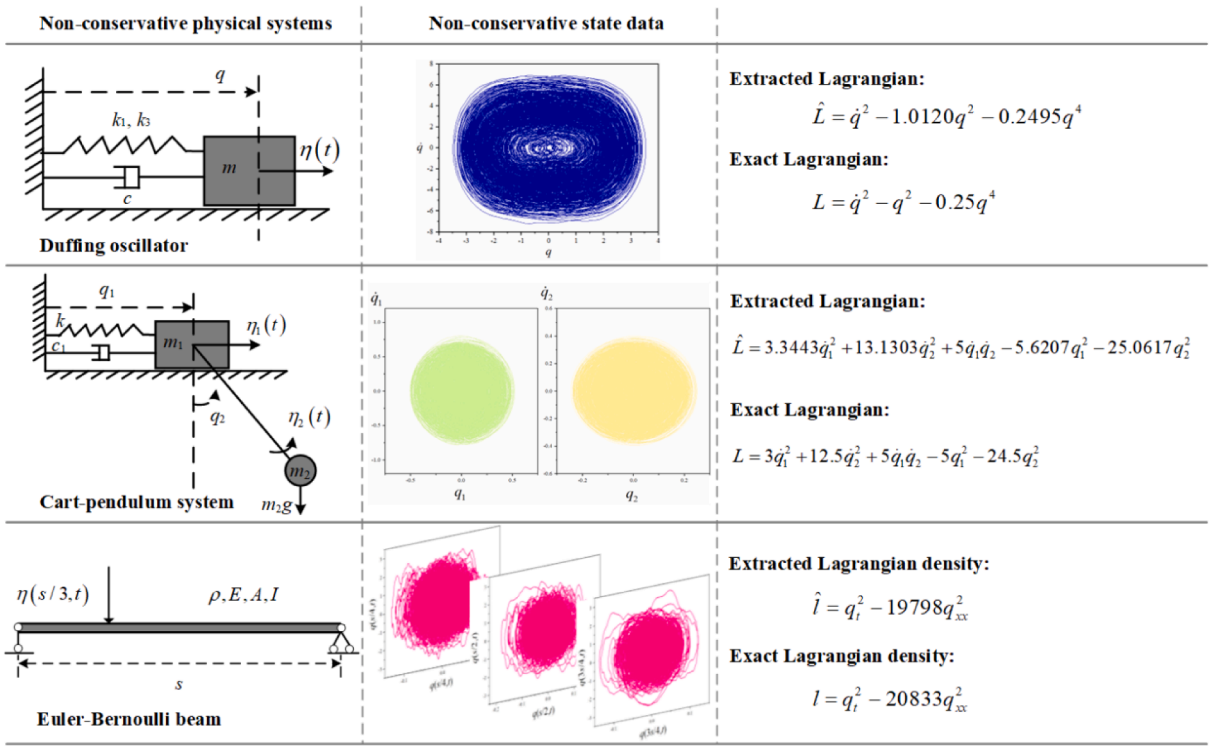

Fig. 3. Summary of identification results. The first row: Duffing oscillator, the second row: cart-pendulum system, and the third row: Euler-Bernoulli beam. The first column: diagrams of physical systems concerned, the second column: nonconservative state data as the starting point, the third column: the extracted Lagrangians and the exact Lagrangians.

fitting accuracy. Consequently, the identified Lagrangian $\widehat{L}(\mathbf{q}, \dot{\mathbf{q}}; \widehat{\alpha})$ is given by Eq. (4) with the $\boldsymbol{\alpha}$ substituted by the optimal solution $\widehat{\alpha}$ in Eq. (8).

3. Case studies

This section is devoted to case studies. We apply the presented method to some illustrative examples, including a Duffing oscillator (single-degree-of-freedom system), a cart-pendulum system (two-degrees-of-freedom system), and a Euler-Bernoulli beam (continuous system). For multi-degrees-of-freedom systems and continuous systems, some complementary discussions are provided. It should be noted that all these three cases have known Lagrangian, such that once the nonconservative forces and system parameters are prescribed, state data can be obtained, which in turn, will be used to extract conservative equations using the present method.

Case 1: Duffing oscillator

Duffing oscillator can be graphically represented by a mechanical system as shown in the first row and first column of Fig. 3. The displacement and velocity of the mass block m are denoted by q and $\dot{q}$, respectively; the linear and cubic nonlinear stiffness are denoted by k_1 and k_3, respectively; $\eta(t)$ is an external excitation that injects energy, while $c(\dot{q})$ is a dissipative component. Together, they induce the nonconservative generalized force $Q(t)$, that is, $Q(t) = \eta(t) - c(\dot{q})$. The conservative structure hidden in this nonconservative system is known and its Lagrangian is $L^*(q, \dot{q}) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}k_1q^2 - \frac{1}{4}k_3q^4$.

In this case study and others to be presented, what is known to us is just the state data $(q,\dot{q})$ in discrete time instants captured from this system; all other information, including mass, stiffness, damping, and excitations, are totally unknown, neither their types nor values. Our objective is to extract the explicit expression of the Lagrangian only from the provided discrete state data. Let us start from a case with random excitation and linear damping, i.e., $\eta(t)$ is Gaussian white noise with intensity $2D$ while $c(\dot{q}) = c\dot{q}$ with linear damping coefficient c.

In step 1, set $m = 1$ kg, k_1=1 N/m, k_3=0.5 N/m^3, $c = 0.01$ Ns/m and $2D$=0.1 N^{2}s. With initial conditions $q(0) = 1$m, $\dot{q}(0) = 0$m /s and a constant time step 0.02 s, numerical integration by the fourth-order Runge-Kutta method produces discrete state data. The state data is collected in the time interval [0, 2000] s (100,000 sampling points in total), as shown in the first row and second column of Fig. 3.

In step 2, the Lagrangian is constructed by a linear combination of low-order polynomials of system states, i.e.,

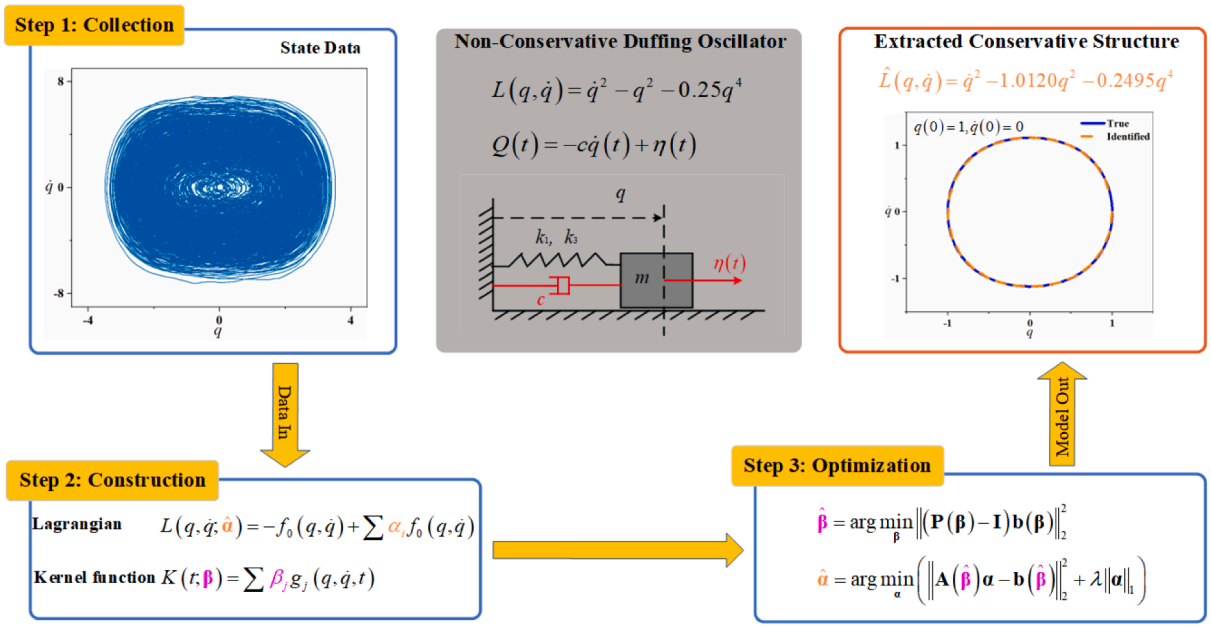

Fig. 4. Flowchart for identifying the conservative structure of randomly excited, dissipative Duffing oscillator. Step 1: collects nonconservative state data, Step 2: constructs Lagrangian and kernel function, and Step 3: optimizes the to-be-determined coefficients of Lagrangian and kernel function.

$$L(q,\dot{q};\boldsymbol{\alpha}) = -\dot{q}^2+\alpha_1\dot{q}^2q+\alpha_2\dot{q}^2q^2+\alpha_3\dot{q}^2q^3+\alpha_4\dot{q}^2q^4+\alpha_5\dot{q}^2q^5 \\ +\alpha_6\dot{q}^2q^6+\alpha_7q^2+\alpha_8q^3+\alpha_9q^4+\alpha_{10}q^5+\alpha_{11}q^6 \tag{9}$$

Note, such terms as $\dot{q}q^k$ ($k = 0, 1, \ldots$) are excluded beforehand because of their triviality; refer to the discussions in Chu and Hayashibe (2020), Hills et al. (2015) and Landau and Lifshitz (1976). Notice that here we start from low-order polynomials; if the loss function of the identified result is small enough, we accept the result; otherwise, we increase the order of polynomials to enlarge the library of basis functions, and then identify again. The kernel function is similarly constructed as:

$$K(t;\boldsymbol{\beta}) = \beta_1q+\beta_2q^2+\beta_3q^3+\beta_4\dot{q}+\beta_5\dot{q}^2+\beta_6\dot{q}^3 \tag{10}$$

in which α_k ($k = 1, \ldots, 11$) and β_k ($k = 1, \ldots, 6$) are coefficients to be determined. Each basis function in Eq. (10) is normalized to have standard deviation 1.

In step 3, select 199 time intervals $[t_0^j, t_1^j]$ from [0, 2000] s with $t_0^j = (j-1)\times 10$s and $t_1^j = j\times 10$ s; each time interval contains 501 sampling points. Derive the 199×11 matrix $\mathbf{A}(\boldsymbol{\beta})$ and 199×1 vector $\mathbf{b}(\boldsymbol{\beta})$ via composite Simpson's rule; then divide $\mathbf{A}(\boldsymbol{\beta})$ and $\mathbf{b}(\boldsymbol{\beta})$ by the standard deviation of $\mathbf{b}(\boldsymbol{\beta})$. We solve the nonlinear optimization problem in Eq. (7) using the genetic algorithm in MATLAB software (MATLAB, 2021). Here, we restrict the search region of each β_k in $[-1, 1]$ for reducing computation complexity. After the optimal solution $\widehat{\beta}$ is found, we solve the regularized least-squares problem in Eq. (8) with $\lambda = 0.1$ to find the optimal $\widehat{\alpha}$.

The final solutions we obtain are, respectively (see Supporting Information for convergence plot of the genetic algorithm),

$$K(t) = 0.2124q-0.3076q^2-0.9233q^3+0.0066\dot{q}+0.0413\dot{q}^2-0.0271\dot{q}^3 \tag{11}$$

$$\widehat{L} = \dot{q}^2-1.0120q^2-0.2495q^4 \tag{12}$$

It is obvious that the sparse Lagrangian identified in Eq. (12) agrees very well with the exact one, that is $L(q,\dot{q}) = \frac{1}{2}\dot{q}^2-\frac{1}{2}q^2-\frac{1}{8}q^4$, except for a constant factor 2 (Landau and Lifshitz, 1976). The identification procedure is summarized and depicted in Fig. 4.

To verify the efficacy of this method to various types of excitations and dissipative components, we summarize the solutions for different combinations in Fig. 5. The first three rows present the combination of linear damping and harmonic excitation, combination of linear damping, random and harmonic excitations, and combination of Coulomb friction and random excitation, respectively. The fourth row presents the case with no excitation and no dissipative component for comparison. For all rows, the first column presents the nonconservative generalized forces unknown to us, the second column depicts the nonconservative state data known to us, while the third column represents the Lagrangian extracted. Obviously, the conservative structure extracted from the nonconservative state data possesses high accuracy for various mechanisms of excitation and dissipation. Parametric analysis indicates that this method is consistently effective for cases with not too large excitation and too large damping, as shown in the left column of Fig. 6, in which the identification error is defined as the average of the relative error of each identified coefficient.

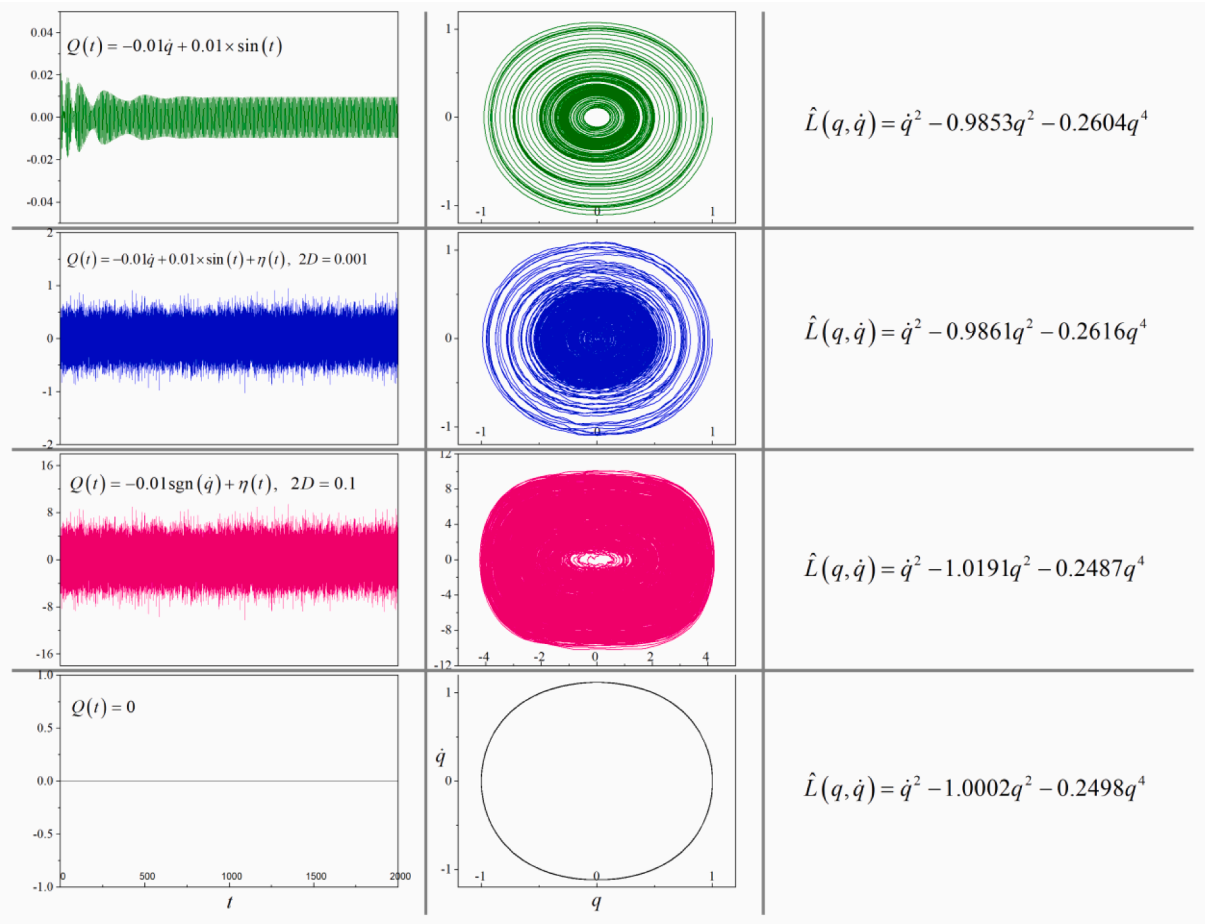

Fig. 5. About Duffing oscillator, the summary of identification results for various nonconservative forces. The first row: linear damping plus harmonic excitation, the second row: linear damping plus harmonic excitation plus random excitation with intensity $2D$=0.001 N^2s, the third row: Coulomb friction plus random excitation with intensity $2D$=0.1 N^2s, and the fourth row: no excitation and no dissipative component. The first column: nonconservative generalized forces; the second column: nonconservative state data, and the third column: the extracted Lagrangian. The exact Lagrangian is $L(q,\dot{q}) = \frac{1}{2}\dot{q}^2 - \frac{1}{2}q^2 - \frac{1}{4}q^4$. The associated kernel functions for these four cases are, respectively, as $K(t) = -0.2991q - 0.4362q^2 + 0.7063q^3 + 0.0971\dot{q} + 0.6125\dot{q}^2 - 0.0656\dot{q}^3$, $K(t) = -0.1353q + 0.2429q^2 + 0.9828q^3 + 0.0187\dot{q} - 0.1407\dot{q}^2 - 0.0166\dot{q}^3$, $K(t) = 0.2472q + 0.1925q^2 - 0.8637q^3 + 0.0242\dot{q} - 0.0215\dot{q}^2 - 0.0457\dot{q}^3$, $K(t) = -0.1086q + 0.6672q^2 - 0.9519q^3 + 0.9985\dot{q} + 1.000\dot{q}^2 + 0.9937\dot{q}^3$.

Case 2: Cart-pendulum system

Herein, we first emphasize some differences encountered in the cases of multi-degree-of-freedom systems. Note that we implement the identification for each degree of freedom separately. Thus, it is quite likely that we will obtain a set of different expressions for the sole Lagrangian. For instance, in the identification associated with the sth degree of freedom, the terms in Lagrangian that do not depend on $(q_s, \dot{q}_s)$ cannot be identified successfully. The expressions identified, denoted by $L_1, L_2, \cdots$, which do not possess any identical term, suggest the separability of the system concerned. The exact Lagrangian will be the sum of all these expressions, that is, $L = L_1 + L_2 + \cdots$. If the identified expressions possess at least one identical term, e.g., $\dot{q}_1\dot{q}_2$ in $L_1 = \dot{q}_1^2 + 2\dot{q}_1\dot{q}_2$ and $L_2 = \dot{q}_2 + \dot{q}_1\dot{q}_2$, we can normalize each expression such that the identical term (e.g., $\dot{q}_1\dot{q}_2$) in each expression has the same coefficient. We then combine all the other terms together to produce the "entire" Lagrangian of the system.

Consider a typical nonlinear two-degree-of-freedom system, a cart-pendulum system, as depicted in the second row and first column of Fig. 3. The linear displacement of the cart with mass m_1 and the angular displacement of the pendulum with mass m_2 are denoted by q_1 and q_2, respectively; the linear stiffness coefficient of the horizontally displaced spring, the pendulum length, and the gravitational acceleration are denoted by k, l, and g, respectively. The nonconservative factors include external excitations (a time-varying force $\eta_1(t)$ imposed on the cart and a time-varying couple $\eta_2(t)$ imposed on the pendulum) and external dissipative components with linear damping coefficients c_1, c_2 (c_2 not shown). The nonconservative generalized forces are $Q_1(t) = -c_1\dot{q}_1 + \eta_1(t)$ and $Q_2(t) = -c\dot{q}_2 + \eta_2(t)$, respectively. The true Lagrangian is $L(\mathbf{q},\dot{\mathbf{q}}) = \frac{1}{2}(m_1 + m_2)\dot{q}_1^2 + \frac{1}{2}m_2l^2\dot{q}_2^2 + m_2l\dot{q}_1\dot{q}_2\cos q_2 - \frac{1}{2}kq_1^2 + m_2gl\cos q_2$.

Herein, $\eta_1(t)$ and $\eta_2(t)$ are described by Gaussian white noise with intensities $2D_1, 2D_2$, and the parameter values are set as m_1=5 kg,

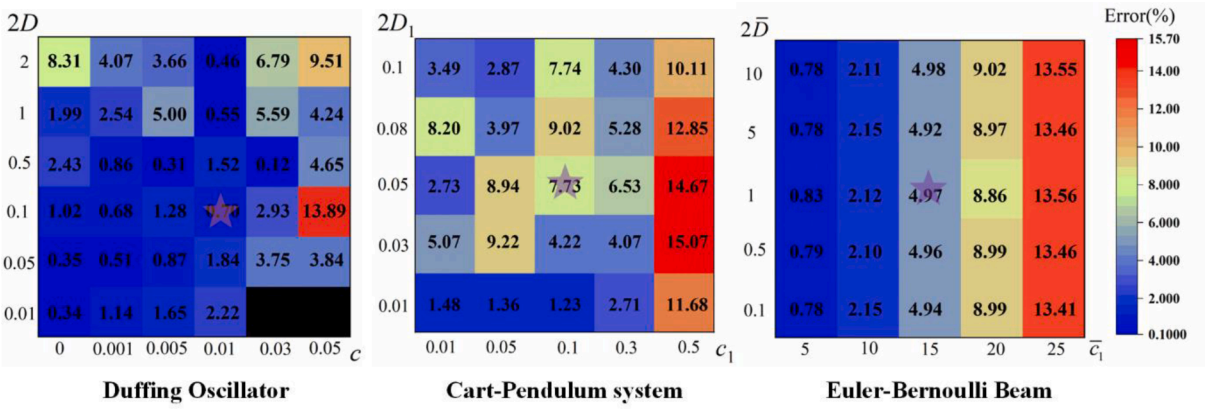

Fig. 6. Identification error versus nonconservative factors. The left one: damping coefficient and excitation intensity for Duffing oscillator; the middle one: damping coefficient and excitation intensity of the linear coordinate for cart-pendulum system with fixed c_2=0.01 and $2D_2$=0.01; and the right one: scaled external damping coefficient and scaled excitation intensity for Euler-Bernoulli beam with fixed scaled internal damping coefficient $\bar{c}_2$=0.001. The purple star denotes the parameters used in the preceding numerical cases. The black square denotes that the identification result does not have the same terms as the true Lagrangian.

m_2=1 kg, k=20 N/m, l=2 m, g=9.8 m/s^2, c_1=0.1 Ns/m, c_2=0.01 Nms, $2D_1$=0.05 N^{2}s, and $2D_2$=0.01N^{2}m^{2}s. In step 1, we numerically integrate the Lagrangian equations by the fourth-order Runge-Kutta method with a fixed time step 0.02 s and initial conditions $q_1(0) = 1\text{m}$, $q_2(0) = 0$, $\dot{q}_1(0) = 0.5\text{m/s}$, and $\dot{q}_2(0) = 0$. We then collect state data in [0, 4000] s (totally 200, 000 sampling points), as shown in the second row and second column of Fig. 3. In step 2, we construct the to-be-identified Lagrangian by a linear combination of given basis functions, that is,

$$\begin{aligned} L = {} & \alpha_1\dot{q}_1^2 + \alpha_2\dot{q}_2^2 + \alpha_3\dot{q}_1\dot{q}_2 + \alpha_4\dot{q}_1\dot{q}_2q_1 + \alpha_5\dot{q}_1\dot{q}_2q_2 + \alpha_6\dot{q}_1\dot{q}_2q_1^2 + \alpha_7\dot{q}_1\dot{q}_2q_2^2 \\ & + \alpha_8\dot{q}_1\dot{q}_2q_1q_2 + \alpha_9\dot{q}_1\dot{q}_2q_1^3 + \alpha_{10}\dot{q}_1\dot{q}_2q_2^3 + \alpha_{11}\dot{q}_1\dot{q}_2q_1^4 + \alpha_{12}\dot{q}_1\dot{q}_2q_2^4 + \alpha_{13}q_1^2 \\ & + \alpha_{14}q_1^3 + \alpha_{15}q_1^4 + \alpha_{16}q_1^5 + \alpha_{17}q_2^2 + \alpha_{18}q_2^3 + \alpha_{19}q_2^4 + \alpha_{20}q_2^5 \end{aligned} \tag{13}$$

Here, the trivial terms, such as $\dot{q}_1q_1^k$ and $\dot{q}_2q_2^k$ with k being non-negative integers, are excluded beforehand. Similarly, we construct the kernel functions as below,

$$W_1(t) = \beta_1^1q_1 + \beta_2^1q_1^2 + \beta_3^1q_1^3 + \beta_4^1\dot{q}_1 + \beta_5^1\dot{q}_1^2 + \beta_6^1\dot{q}_1^3 \tag{14}$$

$$W_2(t) = \beta_1^2q_2 + \beta_2^2q_2^2 + \beta_3^2q_2^3 + \beta_4^2\dot{q}_2 + \beta_5^2\dot{q}_2^2 + \beta_6^2\dot{q}_2^3 \tag{15}$$

In step 3, we select 1999 time intervals $[t_0^j, t_1^j]$ from [0, 4000] s with $t_0^j = (j-1)\times 2\text{s}$ and $t_1^j = j\times 2\text{s}$, and each time interval contains 101 sampling points. We determine the optimal parameter values of $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$ separately. Following similar steps as presented in case 1, the optimal kernel functions found by the nested optimization are, respectively (see Supporting Information for convergence plot of the genetic algorithm),

$$K_1(t) = -0.3175q_1 + 0.1071q_1^2 + 0.7384q_1^3 + 0.0156\dot{q}_1 - 0.0015\dot{q}_1^2 - 0.0424\dot{q}_1^3$$

$$K_2(t) = 0.3946q_2 + 0.0440q_2^2 - 0.8754q_2^3 + 0.0274\dot{q}_2 + 0.0060\dot{q}_2^2 - 0.0510\dot{q}_2^3$$

while the sparse expressions of the associated Lagrangians are

$$L_1 = \dot{q}_1^2 + 1.4951\dot{q}_1\dot{q}_2 - 1.6807q_1^2, L_2 = \dot{q}_2^2 + 0.3808\dot{q}_1\dot{q}_2 - 1.9087q_2^2$$

These two expressions possess one identical term, $\dot{q}_1\dot{q}_2$; thus, we normalize each expression such that the term $\dot{q}_1\dot{q}_2$ in each expression has the same coefficient 5. We then combine all the other terms to produce the "entire" Lagrangian,

$$L = 3.3443\dot{q}_1^2 + 13.1303\dot{q}_2^2 + 5\dot{q}_1\dot{q}_2 - 5.6207q_1^2 - 25.0617q_2^2 \tag{16}$$

To compare with the true Lagrangian, i.e., $L(\mathbf{q},\dot{\mathbf{q}}) = \frac{1}{2}(m_1+m_2)\dot{q}_1^2 + \frac{1}{2}m_2l^2\dot{q}_2^2 + m_2l\dot{q}_1\dot{q}_2\cos q_2 - \frac{1}{2}kq_1^2 + m_2gl\cos q_2$ with the extracted one in Eq. (16), we first need to simplify the expression by keeping the terms no more than quadratic ones—such that $\cos q_2$ in the term $m_2l\dot{q}_1\dot{q}_2\cos q_2$ is taken by 1, and the same function in the term $m_2gl\cos q_2$ by $1 - 1/2q_2^2$. Then the true Lagrangian is approximated as $L = 3\dot{q}_1^2 + 12.5\dot{q}_2^2 + 5\dot{q}_1\dot{q}_2 - 5q_1^2 - 24.5q_2^2 + 49$. Obviously, the extracted Lagrangian in Eq. (16) agrees very well with the approximate Lagrangian, except for a constant 49, a trivial term that is insignificant (Landau and Lifshitz, 1976). Similar to the Duffing oscillator, the identification error almost monotonically increases with increase in excitation intensities and external damping, as shown in the middle column of Fig. 6.

Case 3: Euler-Bernoulli beam

This method can be readily generalized to continuous systems. Here, for simplicity, we illustrate this generalization by one-dimensional continuous systems. The dynamical behavior of a one-dimensional continuous system is described by the following generalized-form Euler-Lagrange equation,

$$\mathrm{EL}\left[l\left(q,\frac{\partial q}{\partial t},\frac{\partial q}{\partial x},\frac{\partial^2 q}{\partial t\partial x}\cdots,\frac{\partial^J q}{\partial t^J},\frac{\partial^J q}{\partial x^J}\right)\right]=Q(x,t),0<x<s \tag{17}$$

where the function $q(x,t)$ of position x and time t denotes the generalized coordinate, while $Q(x,t)$ denotes the nonconservative generalized force distributed in the one-dimensional domain $(0,s)$. l is the Lagrangian density, a scalar function of q and its various-order partial derivatives (assumed up to the J^{th}-order). EL[·] is the Euler-Lagrangian operator for this continuous system, that is,

$$\mathrm{EL}[\,\cdot\,]=\sum_{j=0}^{J}\sum_{k_j=0}^{j}(-1)^j\frac{\partial^j}{\partial x^{k_j}\partial t^{j-k_j}}\left(\frac{\partial[\,\cdot\,]}{\partial\left(\partial^j q/\partial x^{k_j}\partial t^{j-k_j}\right)}\right) \tag{18}$$

For derivation of Eq. (18) from the Hamilton's principle, see Supporting Information.

We embed the generalized-form Euler-Lagrange Eq. (17) in the continuous system concerned. The Lagrangian density replaces the role of the Lagrangian in discrete cases. In contrast to discrete systems, the generalized-form Euler-Lagrange Eq. (17) is a partial differential equation and holds at any spatial position in the domain $(0,s)$. Thus, it is possible to consider only one or several positions and reduce the identification of the continuous system to that of a discrete system, as shown next.

Consider a Euler-Bernoulli beam with simply supported boundaries, as depicted in the third row and first column of Fig. 3. The generalized coordinate $q(x,t)$ describes the deflection of the beam. Denote mass density, elastic modulus, cross-sectional area, moment of inertia, length of the beam, the external and internal damping coefficients by ρ, E, A, I, s, c_1, and c_2, respectively. External excitation $\eta(t)$ is a Gaussian white noise with intensity $2D$ and impressed at position x_e.The exact Lagrangian density to be identified is of the analytical form $l(q_t,q_{xx})=\rho Aq_t^2/2-EIq_{xx}^2/2$; the nonconservative force is $Q(x,t)=-c_1q_t-c_2q_{xxxxt}+\eta(t)\delta(x-x_e)$. Set parameter values as ρ=8000 kg/m^3, $E=200$ GPa, $s=10$ m, A=0.01 m^2, I=8.33 × 10^{-6} m^4 (a square section), x_e=s/3, the scaled external damping coefficient $\bar{c}_1=c_1/\rho A=15$, the scaled internal damping coefficient $\bar{c}_2=c_2/EI=0.001$ and the scaled excitation intensity $2\bar{D}$=$2D\times(2\sin(\pi x_e/s)/\rho As)^2$=1.

In step 1, we combine the modal expansion method (with the first five modals) and the fourth-order Rung-Kutta method (with a fixed time step 0.0001 s and zero initial conditions) to produce spatial-temporal discrete data. We collect the state data in 501 uniform-spaced points, and in [0, 100] s with time step 0.001 s. The phase diagrams of state data at $s/4$, $s/2$ and $3s/4$ are shown in the third row and second column of Fig. 3.

We consider the midpoint by fixing the position $x=s/2$ in the generalized-form Euler-Lagrange Eq. (19), thus reducing the partial differential equation into an ordinary differential form as in the discrete cases. The identification procedures are similar to that of the discrete system. In step 2, construct the to-be-identified Lagrangian by a linear combination of given basis functions, that is,

$$l=\alpha_1q_t^2+\alpha_2q_{xt}^2+\alpha_3q+\alpha_4q^2+\alpha_5q^3+\alpha_6q^4+\alpha_7q_x^2+\alpha_8q_x^3+\alpha_9q_x^4+\alpha_{10}q_{xx}^2+\alpha_{11}q_{xx}^3+\alpha_{12}q_{xx}^4 \tag{19}$$

Similarly, the trivial terms, such as q_tq^k, $q_tq_x^k$, and $q_tq_{xx}^k$ with k being non-negative integers, are excluded. Now, since we focus on the midpoint, we construct the associated kernel function as below,

$$K(t)=\beta_1q\left(\frac{s}{2},t\right)+\beta_2q\left(\frac{s}{2},t\right)^2+\beta_3q\left(\frac{s}{2},t\right)^3+\beta_4q_t\left(\frac{s}{2},t\right)+\beta_5q_t\left(\frac{s}{2},t\right)^2+\beta_6q_t\left(\frac{s}{2},t\right)^3 \tag{20}$$

In step 3, we select 999 time intervals $[t_0^j,t_1^j]$ from [0, 100] s with $t_0^j=(j-1)\times0.1$s and $t_1^j=j\times0.1$s, and each time interval contains 101 sampling points. We apply the Euler-Lagrangian operator in Eq. (18) to the Lagrangian in Eq. (19), fix the position $x=s/2$, multiply by the kernel function in Eq. (20), and then numerically integrate over the selected time intervals to obtain the matrices. The optimal kernel function determined by the nested optimization procedure is as,

$$\begin{aligned}K(t)=&-0.3516q\left(\frac{s}{2},t\right)+0.5845q\left(\frac{s}{2},t\right)^2+0.8430q\left(\frac{s}{2},t\right)^3\\&0.0116q_t\left(\frac{s}{2},t\right)-0.0827q_t\left(\frac{s}{2},t\right)^2-0.0128q_t\left(\frac{s}{2},t\right)^3\end{aligned} \tag{21}$$

and the sparse expression of the Lagrangian density is (see Supporting Information for convergence plot of the genetic algorithm),

$$l=q_t^2-19798q_{xx}^2 \tag{22}$$

For given parameter values, the exact Lagrangian is $l(q_t,q_{xx}^2)=40(q_t^2-20833q_{xx}^2)$. Obviously, the extracted Lagrangian in Eq. (22) is close to the exact Lagrangian except for only a trivial constant factor of 40. As shown in the right column of Fig. 6, the identification error monotonically increases with an increase in external damping, but almost does not depend on the excitation intensity due to the linear property of this dynamical system.

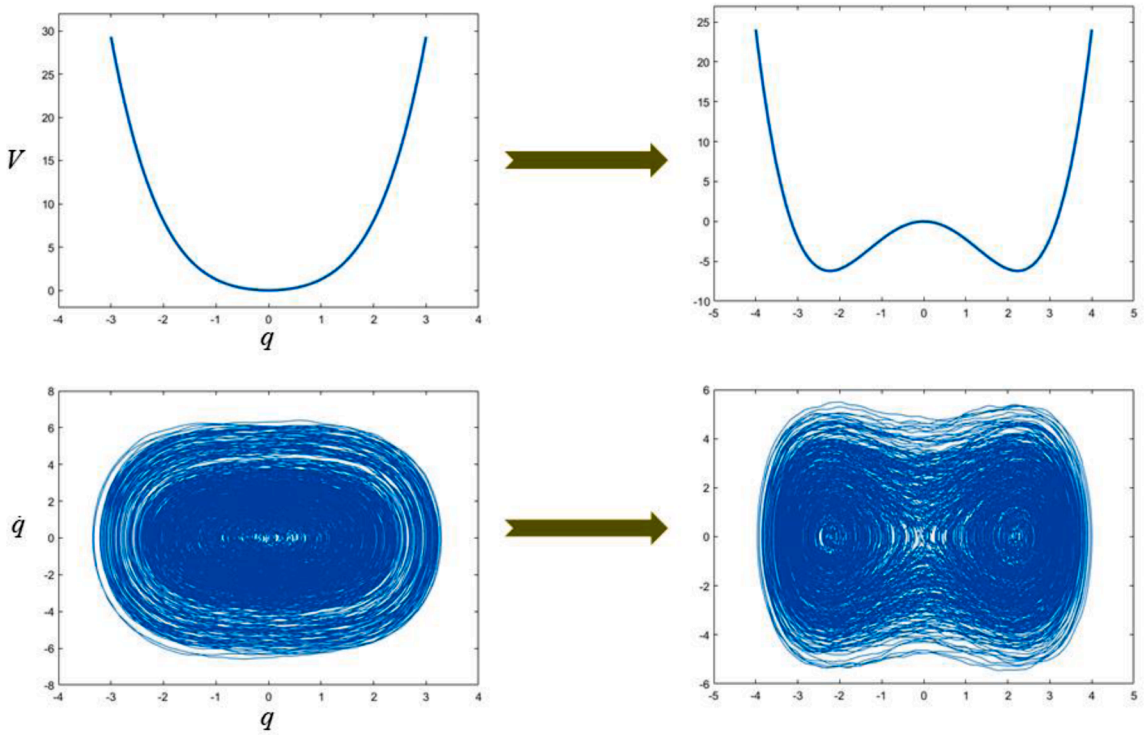

Fig. 7. Potential energy (first row) and phase diagram (second row) of the original (first column) and modified (second column) system for a Duffing oscillator.

4. Conclusions and discussion

This work is devoted to extracting the explicit conservative equations hidden in randomly/deterministically excited, dissipative systems solely from nonconservative state data. In the framework of physics-guided data science, we embed the Euler-Lagrangian equation, the most general description of physical systems, into one unknown system, and extract the Lagrangian (for a discrete system) or Lagrangian density (for a continuous system) based on the concept of orthogonality. The proposed method comes down to an optimization procedure and can successfully extract the conservative structures for three typical dynamical systems.

Once the Lagrangian is discovered, the potential energy is identified at the same time; hence one can employ feedback control to change the qualitative properties of the dynamical system, such as the number and stability of equilibrium points. We illustrate the procedure by the Duffing oscillator examined in Section 3. Using the proposed method, we identify the Lagrangian as $\widehat{L} = \dot{q}^2 - 1.0120q^2 - 0.2495q^4$, from which the potential energy is recognized to be $\widehat{V} = 1.0120q^2 + 0.2495q^4$. By applying a state-dependent feedback control force $F = 7q$, the potential energy of the system is changed to $\widetilde{V} = 1.0120q^2 - \frac{1}{2}\times 7q^2 + 0.2495q^4 = -2.4880q^2 + 0.2495q^4$, as shown in the first row of Fig. 7. As shown in the second row of Fig. 7, the original system oscillates around one equilibrium point in the phase space, i.e., (0, 0); while the modified system with feedback control force oscillates around two equilibrium points.

Intuitively, along this train of thought, to extract the intrinsically conservative structure, we can also embed the integral-variational law or Hamiltonian equations rather than the Euler-Lagrange equations. For the integral-variational law, the arbitrarily selected variations of generalized coordinates serve as kernel functions, as adopted here. The variational functions can be similarly optimized by minimizing the residual. Of course, to satisfy the requirement on variations(i.e., the values of variations must vanish at the time-integral bounds), we can deliberately introduce window functions, which abruptly decrease to zero at the time-integral bounds while almost keep constant somewhere else; and multiply the optimized variations by window functions to guarantee the requirement. For the Hamiltonian equations, however, we must extract the Hamiltonian and the generalized momenta simultaneously. Moreover, the kernel functions and variational functions can also be constructed using neural networks to enlarge the searching space. These will be the subject of our future work.

Author statement

This paper is an original work in this exciting direction, and has not been submitted to other journal for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

YW acknowledges the National Natural Science Foundation of China under Grant Nos. 11872328 and 12132013. YW would like to thank Prof. XL Jin, and Prof. ZL Huang from Zhejiang University for helpful discussions. H.J. acknowledges support from Westlake University.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmps.2022.105127.

References

Bertalan, T., Dietrich, F., Mezić, I., Kevrekidis, I.G., 2019. On learning Hamiltonian systems from data. Chaos: An Interdiscip. J. Nonlinear Sci. 29, 121107 https://doi.org/10.1063/1.5128231.

Brunton, S., Proctor, J., Kutz, J., 2015. Discovering governing equations from data: sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. 113, 3932–3937. https://doi.org/10.1073/pnas.1517384113.

Chattopadhyay, A., Hassanzadeh, P., Pasha, S., 2020. Predicting clustered weather patterns: a test case for applications of convolutional neural networks to spatio-temporal climate data. Sci. Rep. 10, 1317. https://doi.org/10.1038/s41598-020-57897-9.

Chen, T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.K., 2018. Neural ordinary differential equations. Adv. Neural Inf. Process. Syst. 31, 6571–6583. https://doi.org/10.48550/arXiv.1806.07366.

Chen, Z., Liu, Y., Sun, H., 2021. Physics-informed learning of governing equations from scarce data. Nat. Commun. 12, 6136. https://doi.org/10.1038/s41467-021-26434-1.

Choudhary, A., Lindner, J.F., Holliday, E.G., Miller, S.T., Sinha, S., Ditto, W.L., 2021. Forecasting Hamiltonian dynamics without canonical coordinates. Nonlinear Dyn. 103, 1553–1562. https://doi.org/10.1007/s11071-020-06185-2.

Chu, H.K., Hayashibe, M., 2020. Discovering interpretable dynamics by sparsity promotion on energy and the Lagrangian. IEEE Robot. Autom. Lett. 5, 2154–2160. https://doi.org/10.1109/LRA.2020.2970626.

Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P.W., Spergel, D.N., Ho, S., 2020. Lagrangian neural networks. In: ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations. doi:10.48550/arXiv.2003.04630.

Desai, S.A., Mattheakis, M., Sondak, D., Protopapas, P., Roberts, S.J., 2021. Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems. Phys. Rev. E 104, 34312. https://doi.org/10.1103/PhysRevE.104.034312.

Weinan, E., Han, J., Zhang, L., 2021. Machine-learning-assisted modeling. Phys. Today 74, 36–41. https://doi.org/10.1063/PT.3.4793.

Fasel, U., Kutz, J.N., Brunton, B.W., Brunton, S.L., 2021. Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control. Proc. R. Soc. A Math. Phys. Eng. Sci. 478, 20210904 https://doi.org/10.1098/rspa.2021.0904.

Finzi, M., Wang, K.A., Wilson, A.G., 2020. Simplifying Hamiltonian and Lagrangian neural networks via explicit constraints. Adv. Neural Inf. Process. Syst. 33, 13880–13889. https://doi.org/10.48550/arXiv.2010.13581.

Greydanus, S., Dzamba, M., Yosinski, J., 2019. Hamiltonian neural networks. Adv. Neural Inf. Process. Syst. 32, 15379–15389. https://doi.org/10.48550/arXiv.1906.01563.

Hills, D., Grütter, A., Hudson, J., 2015. An algorithm for discovering Lagrangians automatically from data. PeerJ Comput. Sci. 1, e31. https://doi.org/10.7717/peerj-cs.31.

Hirsh, S.M., Barajas-Solano, D.A., Kutz, J.N., 2022. Sparsifying priors for Bayesian uncertainty quantification in model discovery. R. Soc. Open Sci. 9, 211823 https://doi.org/10.1098/rsos.211823.

Horrocks, J., Bauch, C.T., 2020. Algorithmic discovery of dynamic models from infectious disease data. Sci. Rep. 10, 7061. https://doi.org/10.1038/s41598-020-63877-w.

Huang, Z., Tian, Y., Li, C., Lin, G., Wu, L., Wang, Y., Jiang, H., 2020. Data-driven automated discovery of variational laws hidden in physical systems. J. Mech. Phys. Solids 137, 103871. https://doi.org/10.1016/j.jmps.2020.103871.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, 1st ed. Springer, New York.

Kaiser, E., Kutz, J.N., Brunton, S.L., 2018. Discovering conservation laws from data for control. In: Proceedings of the IEEE Conference on Decision and Control, pp. 6415–6421.

Khaled, N., Pooja, V., Andreas, N., 2022. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. Sci. Adv. 5, eaav7903. https://doi.org/10.1126/sciadv.aav7903.

Landau, L.D., Lifshitz, E.M., 1976. Mechanics, 3rd. ed. Butterworth-Heinemann, Oxford.

Lin, Y.K., Cai, G.Q., 1995. Probabilistic Structural Dynamics: Advanced Theory and Application. McGraw-Hill Inc, New York.

Liu, Z., Wang, B., Meng, Q., Chen, W., Tegmark, M., Liu, T.Y., 2021. Machine-learning nonconservative dynamics for new-physics detection. Phys. Rev. E 104, 55302. https://doi.org/10.1103/PhysRevE.104.055302.

MATLAB, 2021. Release 2021a. The Mathworks Inc, Natick, Massachusetts.

Newton, I., Motte, A., Chittenden, N.W., 1846. Newton's Principia : The Mathematical Principles of Natural Philosophy, The Mathematical Principles of Natural Philosophy. Daniel Adee, New York.

Reinbold, P.A.K., Gurevich, D.R., Grigoriev, R.O., 2020. Using noisy or incomplete data to discover models of spatiotemporal dynamics. Phys. Rev. E 101, 10203. https://doi.org/10.1103/PhysRevE.101.010203.

Rudy, S.H., Brunton, S.L., Proctor, J.L., Kutz, J.N., 2017. Data-driven discovery of partial differential equations. Sci. Adv. 3, e1602614 https://doi.org/10.1126/sciadv.1602614.

Schaeffer, H., McCalla, S.G., 2017. Sparse model selection via integral terms. Phys. Rev. E 96, 23302. https://doi.org/10.1103/PhysRevE.96.023302.

Schmidt, M., Lipson, H., 2009. Distilling free-form natural laws from experimental data. Science 324, 81–85.

Stratonovitch, R.L., 1963. Topics in the Theory of Random Noise. Gordon and Breach Science Publishers, New York.

Zhang, S., Lin, G., 2018. Robust data-driven discovery of governing physical laws with error bars. Proc. R. Soc. A Math. Phys. Eng. Sci. 474, 20180305 https://doi.org/10.1098/rspa.2018.0305.

Zhong, Y., Dey, B., Chakraborty, A., 2020. Dissipative SymODEN: encoding Hamiltonian dynamics with dissipation and control into deep learning. In: ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations. doi:10.48550/arXiv.2002.08860.

Zhong, Y., Dey, B., Chakraborty, A., Desmond Zhong, Y., Dey, B., Chakraborty, A., 2019. Symplectic ODE-Net: learning Hamiltonian dynamics with control. In: 8th International Conference on Learning Representations (ICLR 2020). doi:10.48550/arXiv.1909.12077.