



Sparse identification of quasipotentials via a combined data-driven method

Bo Lin¹ · Pierpaolo Belardinelli²

Received: 17 June 2025 / Revised: 30 September 2025 / Accepted: 20 October 2025 / Published online: 20 January 2026
© The Author(s) 2025

Abstract

The quasipotential function allows for comprehension and prediction of the escape mechanisms from metastable states in nonlinear dynamical systems. This function acts as a natural extension of the potential function for non-gradient systems and it unveils important properties such as the maximum likelihood transition paths, transition rates and expected exit times of the system. Here, we demonstrate how to discover parsimonious equations for the quasipotential directly from data. Leveraging machine learning, we combine two existing data-driven techniques, namely a neural network and a sparse regression algorithm, specifically designed to symbolically describe multistable energy landscapes. First, we employ a vanilla neural network enhanced with a renormalization and rescaling procedure to achieve an orthogonal decomposition of the vector field. Next, we apply symbolic regression to extract the downhill and circulatory components of the decomposition, ensuring consistency with the underlying dynamics. This symbolic reconstruction involves a simultaneous regression that imposes constraints on both the orthogonality condition and the vector field. We implement and benchmark our approach using an archetypal model with a known exact quasipotential, as well as a nanomechanical resonator system. We further demonstrate its applicability to noisy data and to a four-dimensional system. Our model-unbiased analytical forms of the quasipotential is of interest to a wide range of applications aimed at assessing metastability and energy landscapes, serving to parametrically capture the distinctive fingerprint of the fluctuating dynamics.

Keywords Quasipotential · Data-driven identification · Neural-networks · Sparse regression

1 Introduction

Examining the stability of attractors and investigating noise-induced switching between metastable solutions is indispensable for understanding critical phenomena in biophysics, engineering, and ecology [1, 2]. These include processes such as protein folding [3], the kinetics of chemical reactions [4], and the behavior of mechanical systems [5, 6]. A thorough understanding of activated escape from a metastable state provides insight into non-equilibrium phenomena that arise from the interplay between nonlinearities and fluctuations [7–9].

For this reason, numerous theoretical and experimental studies across various fields focus on computing and measuring the energy landscape of a system. The escape dynamics of a silica nanoparticle trapped in a bistable potential is extracted from the decaying autocorrelation of the which-well measurement [10]. However, the measurement itself is not inherently well-suited to infer the experimental potential, so models are employed alongside the data to accurately replicate the landscape across all possible configurations, aiming at predicting the turnover. The (quasi-)energy surface of a micro-electromechanical resonator is reconstructed by Dumont et al. through ringdown measurements, revealing attractors and separatrices of the vector field. A model that accurately describes the underlying physics is then retrieved in the form of a Hamiltonian [11]. Mello and Barrick experimentally obtained the protein energy landscape by measuring the stabilities of folded fragments of a repeat protein. The study of the energy surface as a function of the degree of folding of various parts of the protein reveals an early free energy barrier and suggests preferred low-energy pathways for fold-

✉ Pierpaolo Belardinelli
p.belardinelli@univpm.it

¹ Department of Mathematics, National University of Singapore, Lower Kent Ridge Road, Singapore 119076, Singapore

² Department of Construction, Civil Engineering and Architecture, Polytechnic University of Marche, Via Brecce Bianche, Ancona 60131, Italy

ing [12]. A 1D Ising model is then employed to explore unfolding transitions. The landscape of a single unmodified protein, quantifying its change with temperature, is showcased in Peters et al. [13]. Kramers' theory of transition rates is used to model the dynamics, demonstrating strong agreement with the observed state transitions. The potential energy landscape is able to capture the complex phenomenology of disordered materials. An energy evolution model has been introduced to describe the dynamics of the inherent structure in glassy systems [14]. Evaluating the energy landscape is crucial for understanding the presence of nonlinear processes that can give rise to critical transitions or tipping points [15, 16], which occur in various subsystems (tipping elements) of the Earth's climate system [17]. This brief selection of examples, while not exhaustive (cf. [18–20]), highlights the critical importance of accurate modeling in conjunction with precise measurements. Therefore, substantial effort is dedicated to obtaining a mathematical description of the landscape, which serves as a natural metric for assessing stability [21, 22].

In specific scenarios, the task of determining the representation of the energy landscape coincides with identifying the potential function. In gradient systems, the energy landscape directly corresponds to the potential function, meaning that the features and variations of the landscape, such as wells, barriers, and saddle points, are determined by the values and gradients of the potential function. However, when attempting to model the non-equilibrium statistics of transitions between stable states, it is common to encounter non-gradient systems [23]. In this case, the transition dynamics becomes predictable and is entirely defined by the *quasipotential* that extends the idea of energy functions [24]. As a result, the quasipotential landscape provides an intuitive picture of the fundamental dynamical characteristics of complex systems operating outside of equilibrium [25]. With the quasipotential at hand, valuable asymptotic information can be easily estimated. For instance, a calculation over the energy surface allows for the straightforward determination of the maximum likelihood transition path [26]. Additionally, a precise estimate for the invariant probability distribution near an attractor and the expected escape time from a basin of attraction has been formulated on the basis of the quasipotential topology [27]. The quasipotential holds considerable importance as an informative landscape in stochastic dynamical systems, providing insights into the system's equilibria, the transition pathways between metastable states, and the rates of these transitions.

When trying to compute the quasipotential, one encounters a functional minimization problem that can be solved analytically only in specific special cases. In [23], Cameron proposed a quasi-potential solver based on the ordered upwind method [28, 29], which was inspired by Dijkstra's algorithm [30, 31]. A significant development of the ordered upwind method for solving the Hamilton–Jacobi equation

led to a new family of methods for computing the quasipotential on regular meshes, known as the ordered line integral methods (OLIM) [32]. An extension and upgrade of OLIM, incorporating the midpoint quadrature rule, is the olim3D solver [33], which efficiently computes the quasipotential on regular rectangular meshes in 3D. The minimum action methods have also been developed for computing the quasipotential between two specified points by solving a constrained minimization problem of the Freidlin–Wentzell action functional over the path space [34, 35]. A Ritz discretization is used to minimize the action functional in [36]. The geometric minimum action method [37] provides an efficient way for solving the minimization problem based on the arclength parameterization of the path.

Although these approaches have their merits and numerous applications, we propose an alternative solution that does not require solving the variational formulation based on the Freidlin–Wentzell action functional or the associated Hamilton–Jacobi equation over a spatial mesh. Specifically, we employ data-driven methods and machine learning-based approaches for this purpose. The emerging scaling of these techniques has significantly impacted the field of nonlinear dynamics, showcasing their potential for both interpreting complex systems and developing innovative identification methods [38–40]. In this study, we introduce a combination of data-driven approaches to achieve the identification and a symbolic interpretation of the quasipotential from data.

To map the quasipotential landscape from data, a computational method based on neural networks was proposed in [41]. The method identifies an orthogonal decomposition of the vector field that governs the dynamics, in which the quasipotential can be inferred from the potential component. The idea of utilizing a decomposition of the vector field was also used to compute the generalized potential or invariant distribution for randomly perturbed dynamical systems [42, 43]. In [44], a neural network with automatic differentiation inspired by physics-informed neural network is employed to compute the quasipotential based on the Hamilton–Jacobi equation. Estimating the quasipotential with neural networks provides a powerful tool for strategically calculating the mean exit time in a broad class of stochastic dynamical systems [45].

Here, we leverage the approach in [41] to compute a neural network solution of the quasipotential but improve the approach by incorporating normalization and rescaling steps into the input and output layers of the neural networks, respectively. The neural network solution serves as a preparatory step before integrating a constrained regression approach, which can reinterpret the black-box model by neural networks and yield an analytical model. The quasipotential equation adopts a parsimonious and interpretable form, meaning the model is sparse within the space of all possible functions [46, 47], without being restricted to any specific one. This sparse regression technique builds on the

frame of the sparse identification of nonlinear dynamics (SINDy) introduced by [48]. However, we adapt the optimization process to address the specific challenges of the quasipotential problem, particularly focusing on the identification of an involved target matrix. Our routine includes adjustments to preserve orthogonality while ensuring consistency with the original data. Notably, the combined method requires only observed trajectories of the dynamical system as input. Thus, the primary limitation of the technique lies in the availability of such trajectories, which may come directly from observations. By employing a fully data-driven approach, we derive a symbolic quasipotential function without imposing any predefined model for the energy landscape. The scalability of neural networks and sparse regression ensures that our approach is not limited by system dimensionality, making it readily extensible to high-dimensional systems.

The paper is organized as follows. In Sec. 2, we introduce the background of the quasipotential for non-gradient systems and the two parts of the data-driven method. In Sec. 3, we present two applications of the method. We also test it in the context of noisy dynamics, mimicking a realistic imperfect dataset, and extend its applicability to a four-dimensional system. In Sec. 4, we draw the conclusions. The implementation of the method, along with the presented examples, is available under the MIT License in the GitHub repository <https://github.com/LinBoNUS/SIQ>.

2 Methods

2.1 Expanding on the concept of potential function

We undergo the investigation and interpretation of the process through which a dynamical system evolves in presence of multiple equilibria. The underlying physics of the transient dynamics and attractors is assumed to be described by the state equation

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad (1)$$

where $\dot{\mathbf{x}} := d\mathbf{x}/dt$ represents the rate of change of the state variables $\mathbf{x}$ over time, and $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ denotes the nonlinear vector field of dimension d governing the dynamics and responsible for its multistability. In the particular case in which

$$\mathbf{f}(\mathbf{x}) = -\nabla U(\mathbf{x}), \quad (2)$$

we are in the presence of a gradient system, with U referred to as the “potential”. Thanks to this function, the dynamics of this system can be likened to that of a ball-in-cup setup,

where the surface is defined by the potential. In a more general scenario, *i.e.* non-gradient systems, it is not possible to find U such that Eq. (2) is satisfied. However, the concept of a potential landscape can be extended to non-gradient systems through the Freidlin–Wentzell quasipotential function [24], which plays an analogous role in describing the system’s long-term stochastic behavior. Freidlin–Wentzell theory provides a framework for understanding rare noise-driven transitions between metastable states in systems perturbed by small random noise. Central to this theory is the action functional, which quantifies the likelihood of a stochastic trajectory over a given time interval. Let $\mathbf{x}_0$ be an attractor of the system (1), the quasipotential with respect to $\mathbf{x}_0$ is defined as

$$U(\mathbf{x}) = \inf_{T>0} \inf_{\varphi(t)} \int_0^T |\dot{\varphi}(t) - \mathbf{f}(\varphi(t))|^2 dt, \quad (3)$$

where the infimum is over all time $T > 0$ and absolutely continuous path $\varphi(t)$, $0 \leq t \leq T$ connecting $\mathbf{x}_0$ and $\mathbf{x}$. An alternative characterization for the quasipotential can be derived from a decomposition of the vector field [49],

$$\mathbf{f}(\mathbf{x}) = -\nabla V(\mathbf{x}) + \mathbf{g}(\mathbf{x}), \quad (4)$$

where V is solution of the Hamilton–Jacobi equation

$$\nabla V \cdot \nabla V + \mathbf{f} \cdot \nabla V = 0. \quad (5)$$

Assume that the function V is continuously differentiable in $\mathcal{D} \cup \partial\mathcal{D}$ where $\mathcal{D}$ is a bounded domain in $\mathbb{R}^d$ and attains its strict local minimum at a point $A \in \mathcal{D}$. If a further condition that

$$V(\mathbf{x}) > V(A) \text{ and } \nabla V(\mathbf{x}) \neq 0, \quad \text{for } \mathbf{x} \neq A \quad (6)$$

holds, then the function V is a scalar multiple of the quasipotential U with respect to A in a local region containing A [24, 41]. Specifically, that is $U(\mathbf{x}) = 2V(\mathbf{x}) + C$ in the region $\{\mathbf{x} \in \mathcal{D} \cup \partial\mathcal{D} : V(\mathbf{x}) \leq \min_{\mathbf{y} \in \partial\mathcal{D}} V(\mathbf{y})\}$, where C is a constant.

The quasipotential represents a useful generalization of the potential function for a non-gradient system. In terms of the vector field, the $-\nabla V$ component in Eq. (4) spurs a ball to roll to the bottom of a valley in the ball-in-a-cup analogy, thus V specifies a hyper-surface in which all the trajectories move “downhill” in the absence of perturbations. From Eq. (4) and the Hamilton–Jacobi relation (5) it is straightforward to verify that

$$\mathbf{g} \cdot \nabla V = 0, \quad (7)$$

hence, $\mathbf{g}$ and ∇V are perpendicular. Without additional external forces, the “circulatory” component $\mathbf{g}$ creates the

circulation of trajectory around levels of V . For convenience, we refer to $\nabla V(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ in the decomposition (4), (7) as downhill and circulatory components, respectively [23, 24].

Our method derives V without solving numerically Eq. (5). A two-step data-driven approach is designed to output the symbolic expression of the quasipotential in terms of the state space variables, *i.e.* $V(\mathbf{x})$. In the first part of the method, a neural network determines the elements in the orthogonal decomposition of Eq.(4) (see Sec.2.2). Neural networks are recognized as universal function approximators, capable of approximating any continuous function on a compact domain to arbitrary accuracy, given sufficient capacity and a suitable architecture [50, 51]. This property makes them particularly effective for learning complex, nonlinear relationships between inputs and outputs without requiring an explicit analytical model.

In the second part of the method, the learned mapping functions, which are not associated with a physically interpretable set of equations, are reinterpreted via constrained regression (see Sec. 2.3). This technique identifies concise dynamic models by expressing the vector field as linear combinations of a large library of candidate functions and promoting sparsity in the coefficients through methods such as sequential threshold least squares or the lasso. Popularized by the SINDy framework [48], this approach automatically selects only the essential terms required to reproduce the observed time series, yielding interpretable models that minimize overfitting. By discarding negligible basis functions, sparse regression provides direct insight into the underlying governing equations and remains computationally efficient even for high-dimensional systems.

2.2 Dataset and neural networks

The downhill and circulatory components in the orthogonal decomposition (4) are identified with two single neural networks, namely V_θ and $\mathbf{g}_\theta$. Hence, the parameterized vector field becomes

$$\mathbf{f}_\theta(\mathbf{x}) = -\nabla V_\theta(\mathbf{x}) + \mathbf{g}_\theta(\mathbf{x}). \quad (8)$$

The parameterization is to be identified among a set of trajectories in phase space that describes the dynamics of the unknown system. Specifically, we denote the trajectory data by $\{(X_i^{j,0}, X_i^{j,h})\}$, $1 \leq i \leq N$, $1 \leq j \leq M$, being N the number of trajectories, M the number of data pairs along each trajectory, and $X_i^{j,0}$ and $X_i^{j,h}$ the observed state positions for the i th trajectory at the time t_j and $t_j + h$, respectively.

The networks V_θ and $\mathbf{g}_\theta$ are trained such that the error in the predicted dynamics is minimized and simultaneously to ensure the orthogonality condition $\mathbf{g}_\theta \cdot \nabla V_\theta = 0$. Let $\mathcal{I}_h(X_i^{j,0}; \mathbf{f}_\theta)$ be the state at time h for the system $\dot{\mathbf{x}} = \mathbf{f}_\theta(\mathbf{x})$

integrated by a suitable numerical scheme starting from $X_i^{j,0}$, then the

loss function is taken as

$$L(\theta) = \frac{1}{NM} \sum_{i,j} \left\| \frac{1}{h} (\mathcal{I}_h(X_i^{j,0}; \mathbf{f}_\theta) - X_i^{j,h}) \right\|_2^2 + \frac{\lambda}{S} \sum_{k=1}^S w \left(\frac{\nabla V_\theta(\tilde{X}_k) \cdot \mathbf{g}_\theta(\tilde{X}_k)}{|\nabla V_\theta(\tilde{X}_k)| \cdot |\mathbf{g}_\theta(\tilde{X}_k)|}, \delta \right). \quad (9)$$

In Eq. (9), the piecewise function $w(z, \delta) = z^2 \mathbb{I}_{z \geq 0} + \delta z^2 \mathbb{I}_{z < 0}$ with $\delta = 0.1$, λ is a parameter controlling the strength of the loss term corresponding to the orthogonality condition, and $\{\tilde{X}_k\}$, $1 \leq k \leq S$ is a representative subset of $\{X_i^{j,0}\}$ with a uniform distribution in the state space.

We parameterize the potential component as $V_\theta(\mathbf{x})$ by the sum of a vanilla neural network $V_\theta^{\text{NN}}(\mathbf{x})$ with hyperbolic tangent (tanh) as the activation function and a harmonic function [41]. This parameterization ensures the properties that V_θ is a radially unbounded function, and the set $H = \{\mathbf{x} \in \mathbb{R}^d : \nabla V_\theta(\mathbf{x}) = 0\}$ is a bounded and Lebesgue measure-zero set in $\mathbb{R}^d$. As a subset of H , the set of all minimizers of the function $V_\theta(\mathbf{x})$ is also bounded and has the measure zero. Thus, $V_\theta(\mathbf{x})$ serves as a candidate function with good physical properties for approximating the quasipotential.

In addition, the training performances with the loss function (9) are ameliorated by adding the following two steps in the initialization of the neural networks: *i*) a normalization step to the input layer of $V_\theta, \mathbf{g}_\theta$ with respect to the data points; *ii*) a scalar multiplication to the output of the networks. In details, we construct the two networks V_θ and $\mathbf{g}_\theta$ as

$$V_\theta(\mathbf{x}) = \eta_v \cdot \tilde{V}_\theta \left(\frac{\mathbf{x} - \mu}{\sigma} \right), \quad \text{in which } \tilde{V}_\theta(\mathbf{y}) = V_\theta^{\text{NN}}(\mathbf{y}) + |\mathbf{y}|^2 \quad (10)$$

and

$$\mathbf{g}_\theta(\mathbf{x}) = \eta_g \cdot \mathbf{g}_\theta^{\text{NN}} \left(\frac{\mathbf{x} - \mu}{\sigma} \right), \quad (11)$$

with $\mathbf{g}_\theta^{\text{NN}}$ being a vanilla neural network mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$ and μ, σ being the sample mean and deviation of the data points $\{X_i^{j,0}\}$, respectively. The parameters η_v, η_g rescale the initial magnitudes of $|\nabla V_\theta|, |\mathbf{g}_\theta|$ to that of $|\mathbf{f}|$ estimated from the data. This is obtained by

$$\eta_v = \arg \min_{\eta > 0} \sum_{i,j} \left(|\eta \mathbf{v}_i^j| - |Y_i^j| \right)^2 = \sum_{i,j} |\mathbf{v}_i^j| \cdot |Y_i^j| / \sum_{i,j} |\mathbf{v}_i^j|^2 \quad (12)$$

and

$$\begin{aligned}\eta_g &= \arg \min_{\eta > 0} \sum_{i,j} \left(|\eta \mathbf{g}_i^j| - |Y_i^j| \right)^2 \\ &= \sum_{i,j} |\mathbf{g}_i^j| \cdot |Y_i^j| / \sum_{i,j} |\mathbf{g}_i^j|^2.\end{aligned}\quad (13)$$

Here, $\mathbf{v}_i^j = \nabla \tilde{V}_\theta((X_i^{j,0} - \mu)/\sigma)$, $\mathbf{g}_i^j = \mathbf{g}_\theta^{\text{NN}}((X_i^{j,0} - \mu)/\sigma)$ and $Y_i^j = (X_i^{j,h} - X_i^{j,0})/h$ is the estimated vector for $\mathbf{f}$ at the state $X_i^{j,0}$. Note that the parameters μ , σ , η_v and η_g are kept as constant during the subsequent training process. This procedure has not been implemented in the earlier work [41].

2.3 Sparse symbolic regression of the quasipotential function

Based on the neural network approximations for the potential V and circulatory vector $\mathbf{g}$, We determine a sparse symbolic expression for the orthogonal decomposition of $\mathbf{f}$, which facilitates its physical interpretation and enables analytical treatment. This is achieved by projecting the dynamics over a minimal set of functions from a candidate library $\Theta(\mathbf{X}) = [\theta_1(\mathbf{X}), \dots, \theta_q(\mathbf{X})]$, which is appropriately selected standing for possible functions. For V and $\mathbf{g}$, the symbolic representation reads

$$\begin{aligned}V^{\text{Symb}}(\mathbf{X}) &= \Theta(\mathbf{X}) \mathbf{\Xi}_v, \\ \mathbf{g}^{\text{Symb}}(\mathbf{X}) &= \Theta(\mathbf{X}) \mathbf{\Xi}_g,\end{aligned}\quad (14)$$

where $\mathbf{\Xi}_v$ and $\mathbf{\Xi}_g$ are the vectors of the unknown coefficients to be determined. In addition, $\mathbf{X}$ represents a NM -by- d matrix containing the data points $\{X_i^{j,0}\}$ from the sampled trajectories with multiple initial conditions. The data domain of the symbolic regression does not necessarily coincides with that of the neural networks. Indeed, it is convenient not to overextend the identification far from the metastable states of the system. Outer boundaries of V are inherently unreliable due to scarcity of data. We limit the selection over $\mathbf{X}$ by defining a subset of $\{X_i^{j,0}\}$ below a potential threshold value τ , i.e. $V_\theta(\mathbf{x}) < \tau$.

For a number of function libraries $\Theta(\mathbf{X})$, e.g. polynomials or trigonometric functions, the gradient of the basis function, $\nabla \theta_k(\mathbf{X})$, can be expressed as a linear combination of the basis functions themselves. In this case, the gradient of the potential in Eq. (14) can still be represented via the same function library Θ , that is $\nabla V^{\text{Symb}}(\mathbf{X}) = \Theta(\mathbf{X}) T(\mathbf{\Xi}_v)$, where $T: \mathbb{R}^{q \times 1} \rightarrow \mathbb{R}^{q \times d}$ is a linear transformation function of the vector $\mathbf{\Xi}_v$. Therefore, the vector field is approximated as

$$\begin{aligned}\mathbf{f}^{\text{Symb}}(\mathbf{X}) &= -\nabla V^{\text{Symb}}(\mathbf{X}) + \mathbf{g}^{\text{Symb}}(\mathbf{X}) \\ &= \Theta(\mathbf{X})(-T(\mathbf{\Xi}_v) + \mathbf{\Xi}_g).\end{aligned}\quad (15)$$

Now, existing sparse regression approaches aim to directly match observable data or byproducts such as their derivatives [52]. Unfortunately, directly inferring a symbolic regression of the quasipotential from data is not feasible. The essential information about the energy landscape remains concealed unless we possess prior knowledge on how to divide the dynamics into downhill and circulatory components. Therefore, we structure the symbolic identification to extract the quasipotential function from the decomposition provided by the neural networks. Specifically, a target matrix for the identification is set as

$$\mathbf{G}(\mathbf{X}) = [\mathbf{f}_\theta(\mathbf{X}), V_\theta(\mathbf{X}), \mathbf{g}_\theta(\mathbf{X})]. \quad (16)$$

We determine the unknown matrix $\mathbf{\Xi} \in \mathbb{R}^{q \times (2d+1)}$ which contains coefficient vectors for $\mathbf{f}^{\text{Symb}}$, V^{Symb} and $\mathbf{g}^{\text{Symb}}$, via the constrained sparse relaxed regularized regression

$$\arg \min_{\mathbf{W}, \mathbf{\Xi}} \frac{1}{2} \|\mathbf{G}(\mathbf{X}) - \Theta(\mathbf{X}) \mathbf{\Xi}\|^2 + \lambda R(\mathbf{W}) + \frac{1}{2\nu} \|\mathbf{\Xi} - \mathbf{W}\|^2, \quad (17)$$

where $R(\cdot)$ is a l_0 -regularization term that promotes sparsity and minimizes over-fitting and $\mathbf{W}$ is an auxiliary variable which is introduced here to enable relaxation and partial minimization in order to improve the conditioning of the problem and tackle the non-convexity of the optimization [53]. In addition, λ and ν are hyper-parameters that control the strength of regularization and relaxation, respectively. As indicated from Eq. (15),

$$\mathbf{\Xi} = [-T(\mathbf{\Xi}_v) + \mathbf{\Xi}_g, \mathbf{\Xi}_v, \mathbf{\Xi}_g] \quad (18)$$

includes relations between the coefficients themselves. These are incorporated in the optimization problem (17) through constraints between the coefficients, denoted by

$$\mathbf{C} \mathbf{\Xi} = \mathbf{0}. \quad (19)$$

Remark In general, the hypothesis space for V or $\mathbf{g}$ may be chosen as the linear span of a subset of the library $\Theta(\mathbf{X})$, rather than utilizing the entire set. For example, one may perform regression in the space of polynomials in $\mathbf{x}$ up to fourth order for V and polynomials up to third order for $\mathbf{g}$. The absence of some terms in a specific library subset can be readily achieved by adding constraining equations to Eq. (14). For instance, specific coefficients of $\mathbf{\Xi}_v$ and $\mathbf{\Xi}_g$ can be set to zero, which is incorporated into the optimization problem (17).

We note that, in addition to the constraint (19), it is possible to incorporate physical knowledge to guide the sparse regression toward a specific representation of the dynamics (see, e.g., [54]). We solve the constrained optimization problem

(17), (19) by performing an initialization step for the coefficient matrix Ξ (see Appendix A), obtaining V^{Symb} . The symbolic expression for the Freidlin-Wentzell quasipotential is its scalar multiple, $U^{\text{Symb}}(\mathbf{x}) = 2 \cdot V^{\text{Symb}}(\mathbf{x})$. Finally, we note that projecting onto a set of continuous functions accurately handles only continuously differentiable energy surface topologies. However, the regression method can be extended to accommodate discontinuities, cf. [55] and references therein.

3 Results

To provide a context for our study and test the efficacy of our machine learning approach, we investigate an archetypal system with known quasipotential (Sec. 3.1). This example is further examined by introducing noise and extended to four dimensions in Sec. 3.1.1 and Sec. 3.1.2, respectively. Subsequently, we employ the combined data-driven technique to analyze a system of engineering interest: the dynamics of a nanomechanical graphene resonator (Sec. 3.2).

In both the applications, we create a synthetic set of data points by sampling trajectories of the system and taking snapshots of the trajectories. In the first part of the method (Sec. 2.2), we train neural networks to reconstruct an orthogonal decomposition of the vector field. Details for sampling the data and training the neural networks are provided in Appendix B. In the second part of the method (Sec. 2.3), we identify sparse symbolic expressions for the downhill and circulatory components that define the dynamics of the system.

3.1 Example 1: An archetypal model with exact quasipotential

First, we consider a non-gradient dynamical system in the three-dimensional space,

$$\begin{cases} \dot{x} = -2(x^3 - x) - y - z, \\ \dot{y} = -y + 2(x^3 - x), \\ \dot{z} = -z + 2(x^3 - x), \end{cases} \quad (20)$$

The system has two stable equilibrium points at $A = (-1, 0, 0)$ and $B = (1, 0, 0)$ and the quasipotential with respect to the two states is given by $U(\mathbf{x}) = x^4 - 2x^2 + y^2 + z^2 + 1$.

We generate 5000 trajectories of the system by numerically simulating the dynamics with initial states sampled from the computational domain $\Omega = [-2, 2] \times [-1.5, 1.5] \times [-1.5, 1.5]$. Figure 1 (a) shows a plot of five generated trajectories. A dataset $\{X_i^{j,0}, X_i^{j,h}\}$ is then constructed by taking snapshots from the generated trajectories. From the data, we

first reconstruct an orthogonal decomposition of the vector field by training the neural networks $V_\theta(\mathbf{x})$ and $\mathbf{g}_\theta(\mathbf{x})$ with the loss function (9). Subsequently, we perform the sparse symbolic regression for the downhill and circulatory components in the decomposition. The identified components are showcased in Fig. 1 (b) and (c). Specifically, we take the basis functions for the regression as polynomials in (x, y, z) up to fifth order for both $V^{\text{Symb}}, \mathbf{g}^{\text{Symb}}$. The target matrix is constructed with the neural network solutions over a subset of $\{X_i^{j,0}\}$ built with a potential threshold $\tau = \min_\Omega V_\theta(\mathbf{x}) + 2$. In the learning problem (17), we take the parameters $\lambda = 0.1$ and $\nu = 10^{-5}$.

The identified quasipotential and downhill/circulatory components, $U^{\text{Symb}}(\mathbf{x})$, $\nabla V^{\text{Symb}}(\mathbf{x})$, $\mathbf{g}^{\text{Symb}}(\mathbf{x})$, as compared against the exact ones, are reported in Table 1. From the table, one can clearly observe that the identified expressions have the same like terms as in the exact ones with an error on the order of 10^{-3} for the corresponding coefficients. The results indicate that the proposed method is able to discover the explicit expression for the quasipotential of nonlinear systems from the observed data. Also, we show contour plots of the exact quasipotential $U(\mathbf{x})$ and identified potential $U^{\text{Symb}}(\mathbf{x})$ in Fig. 1 (d) and (e). The accuracy of the method is demonstrated by the fact that the learned potential U^{Symb} is almost indistinguishable to U .

We implement the proposed method in 50 independent runs. On each run, we generate 5000 trajectories from randomly sampled initial states, train the neural networks and perform the sparse symbolic regression, in the same way as described above. In Fig. 2, we show the distributions of the learned coefficients for each library term in $U^{\text{Symb}}(\mathbf{x})$. The relative root mean square error of the computed coefficients in $U^{\text{Symb}}(\mathbf{x})$ corresponding to the nontrivial terms ($1, x^2, y^2, z^2$ and x^4) is $8.87 \times 10^{-4} \pm 3.48 \times 10^{-4}$. The statistics of the results confirm that the proposed method is able to predict the coefficients of nontrivial terms with high accuracy and identify the trivial terms of the quasipotential (see Fig. 3).

Long-term prediction. We proceed to assess the accuracy of the identified dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$ in predicting the evolution of the system. A set of 1000 trajectories ($\{X_k(t)\}$, $1 \leq k \leq 1000$) not included in the previous ensemble are assessed. For each generated trajectory, we compute the error

$$E_k = \frac{\|X_k^{\text{Symb}}(t) - X_k(t)\|_2}{\|X_k(t)\|_2} \quad (21)$$

in order to quantify the difference between the original and learned dynamics. The statistics of the errors for the 1000 generated trajectories are $3.04 \times 10^{-4} \pm 6.55 \times 10^{-3}$. The comparison of six generated trajectories with the observed ones in Fig. 4 (a) showcases the remarkable accuracy of the method.

Table 1 The exact and identified expressions for the quasipotential $U(\mathbf{x})$, downhill component $\nabla V(\mathbf{x})$ and circulatory component $\mathbf{g}(\mathbf{x})$ in the decomposition of the vector field for the system (20). Three digits after the decimal point are displayed in the identified expressions

	Exact expression	Identified expression
$U(\mathbf{x})$	$x^4 - 2x^2 + y^2 + z^2 + 1$	$1.000x^4 - 2.001x^2 + 0.999y^2 + 0.999z^2 + 1.001$
$\nabla V(\mathbf{x})$	$\begin{bmatrix} 2x^3 - 2x \\ y \\ z \end{bmatrix}$	$\begin{bmatrix} 2.001x^3 - 2.001x \\ 0.999y \\ 0.999z \end{bmatrix}$
$\mathbf{g}(\mathbf{x})$	$\begin{bmatrix} -y - z \\ 2x^3 - 2x \\ 2x^3 - 2x \end{bmatrix}$	$\begin{bmatrix} -1.000y - 1.000z \\ 2.003x^3 - 2.004x \\ 2.003x^3 - 2.003x \end{bmatrix}$

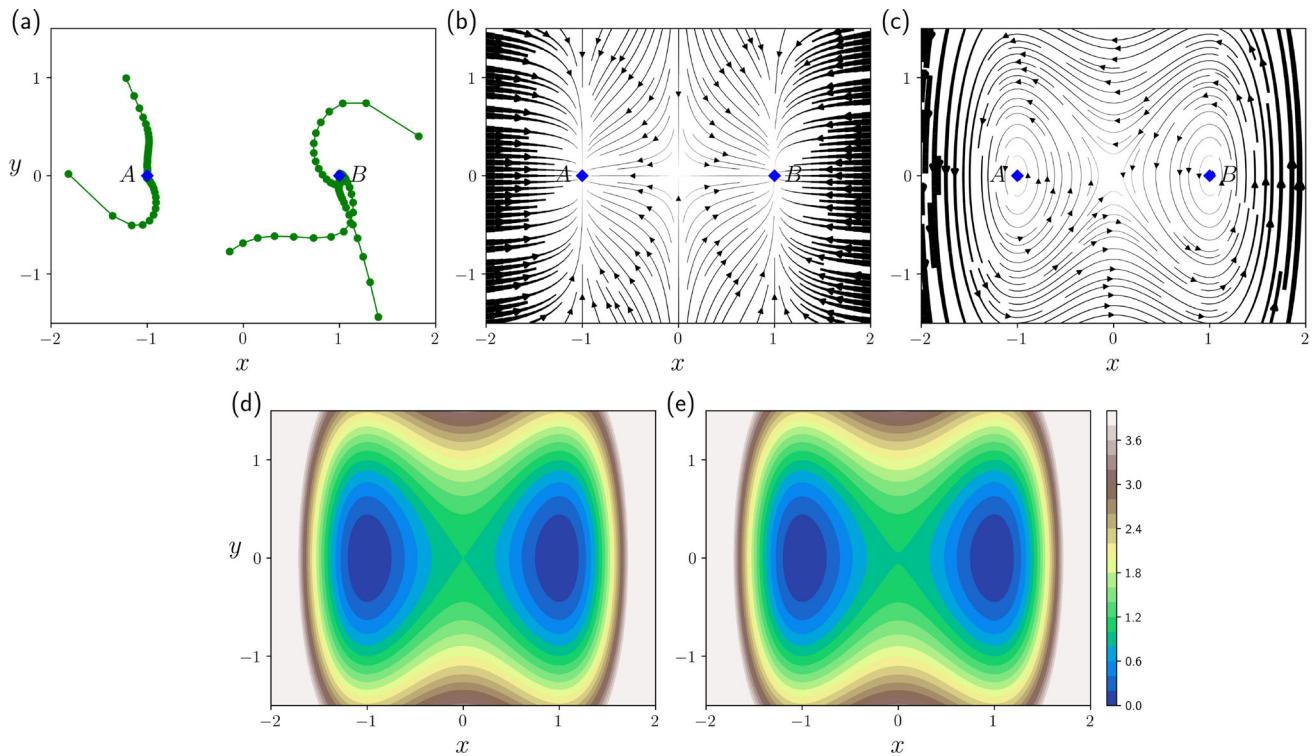


Fig. 1 Combined data-driven method applied to the system (20) with 5000 sampled trajectories. (a) Five generated trajectories. (b) The identified downhill component $-\nabla V^{\text{Symb}}(x, y, z)$. (c) The identified circulatory component $\mathbf{g}^{\text{Symb}}(x, y, z)$. (d) The exact quasipotential

$U(x, y, z)$. (e) The identified quasipotential $U^{\text{Symb}}(x, y, z)$. All plots are projected on the (x, y) -plane. In panels (b) and (c), the line thickness shows the flow velocity

Invariant distribution. With the identified quasipotential $U^{\text{Symb}}(\mathbf{x})$ as in Table 1, one is able to infer explicitly the invariant distribution of the system (20) in the presence of a white noise. The randomly perturbed dynamics is described by the stochastic differential equation:

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t)dt + \sqrt{\epsilon} dW_t, \quad t > 0, \quad (22)$$

where ϵ is a parameter controlling the strength of the noise and W_t is a Wiener process. When ϵ is small, the invariant distribution of the system (22) can be approximated via the

quasipotential $U(\mathbf{x})$ [24, 42, 43],

$$p(\mathbf{x}) = Z^{-1} \exp(-U(\mathbf{x})/\epsilon), \quad (23)$$

where Z is a normalization constant.

For the system (22), computing the invariant distribution $p_\epsilon(\mathbf{x})$ with the identified quasipotential $U^{\text{Symb}}(\mathbf{x})$ allows us to accurately estimate the normalization constant Z . This calculation is detailed in Appendix C. The estimated constants, by accounting for various values of ϵ (0.05, 0.1 and 0.2), are

$$Z_{0.05} = 0.0625; \quad Z_{0.1} = 0.1794; \quad Z_{0.2} = 0.5236. \quad (24)$$

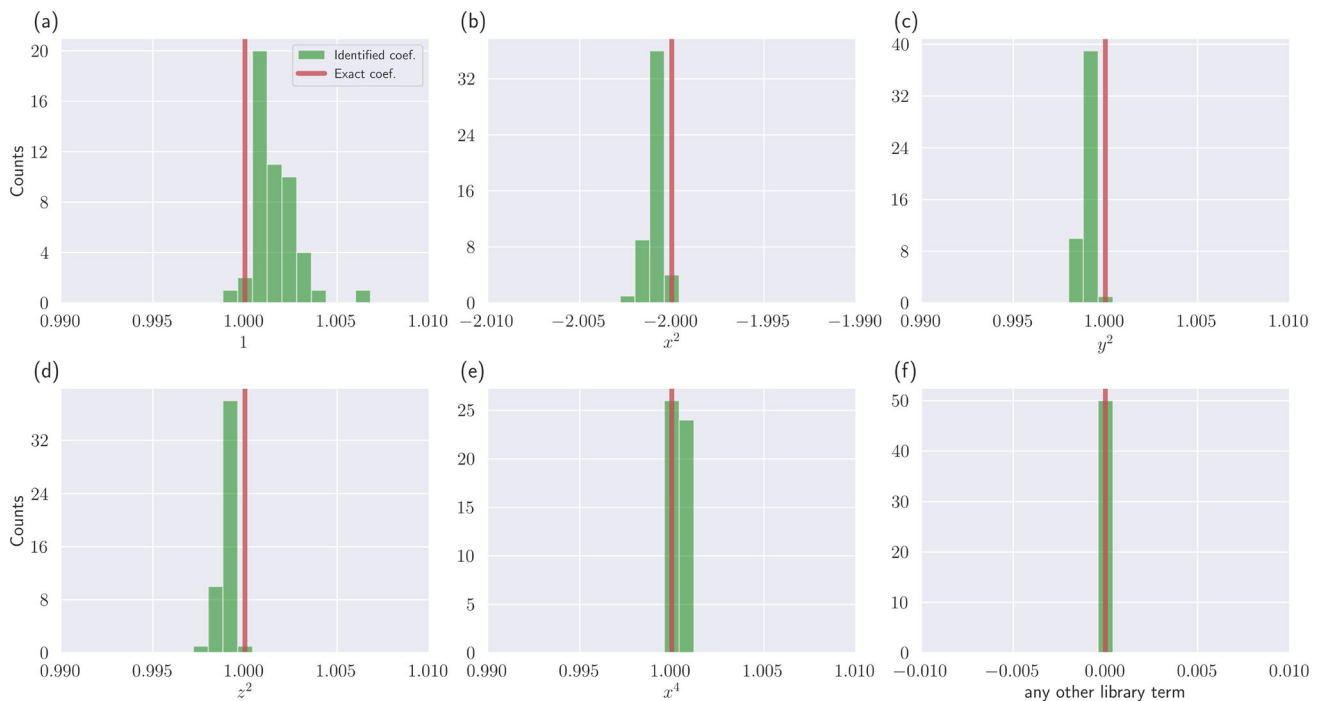


Fig. 2 "Distributions of the coefficients for each library term in the identified potential $U^{\text{Symb}}(\mathbf{x})$ in 50 independent runs for the system in Eq. (20). Panel (a)–(e) indicate the coefficient distributions for the non-

trivial terms ($1, x^2, y^2, z^2$ and x^4) as in $U(\mathbf{x})$, while panel (f) shows that learned coefficients are zero in all runs for any of the remaining library terms

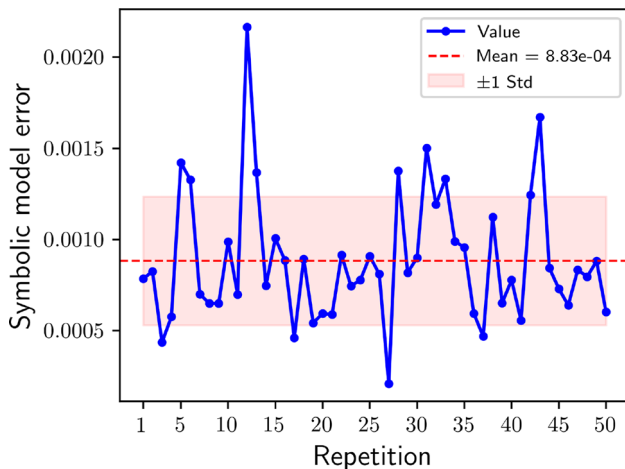


Fig. 3 Error of the symbolic model with respect the exact one for the 50 repetitions of the data-driven symbolic identification

Therefore, the explicit expression of the invariant distribution is given by

$$p_{\epsilon}(\mathbf{x}) = Z_{\epsilon}^{-1} \exp(-U^{\text{Symb}}(\mathbf{x})/\epsilon). \quad (25)$$

Fig. 4 (b)–(d) illustrates the evolution of $p_{\epsilon}(\mathbf{x})$ for different values of ϵ . As the noise intensity increases, the peaks of the invariant distribution broaden, with the probability density spreading out in the (x, y) phase space.

3.1.1 Application of the method to noisy data

A more practical situation is that the system is under the influence of random perturbations. To validate the ability of the method for dealing with such systems, we consider the system in Eq. (20) perturbed by small noise,

$$\begin{cases} \dot{x} = -2(x^3 - x) - y - z + \xi_1(t), \\ \dot{y} = -y + 2(x^3 - x) + \xi_2(t), \\ \dot{z} = -z + 2(x^3 - x) + \xi_3(t), \end{cases} \quad (26)$$

where $\xi(t) = (\xi_1(t), \xi_2(t), \xi_3(t))$ is a white noise with $\langle \xi_j(t) \xi_k(0) \rangle = 2\epsilon \delta_{jk} \delta(t)$ in which ϵ is a parameter specifying the magnitude of the noise. In the example, we take the parameter $\epsilon = 0.01$.

We generate a dataset of 5000 noisy trajectories by numerically simulating the dynamics using the Euler-Maruyama scheme. A plot of the sampled trajectories, as compared against the ones sampled from the deterministic system (20) with the same initial conditions, is shown in Fig. 5 (a). Then, with the noisy trajectories, we implement the method for training the neural networks with the loss function (9) and performing the sparse symbolic regression by solving the problem (17), (19). All the parameters in the data generation, neural network training and symbolic regression are taken as

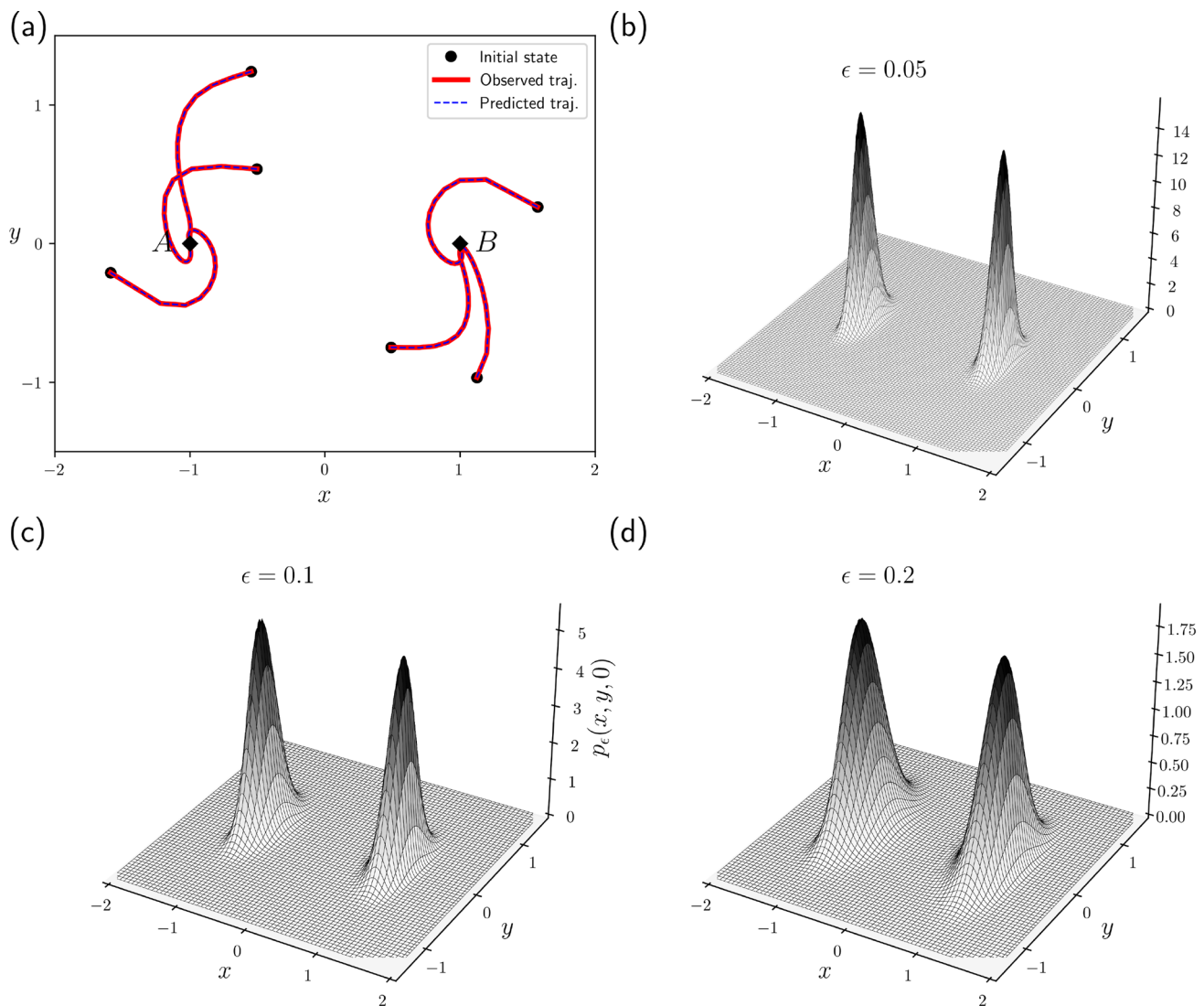


Fig. 4 (a) Phase plot of six observed trajectories and the trajectories generated by simulating the dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$ with the same initial states. (b)-(d) Plots of the invariant distribution computed by

$p_\epsilon(x, y, z) = Z_\epsilon^{-1} \exp(-U^{\text{Symb}}(x, y, z)/\epsilon)$ of the randomly perturbed dynamics system (22) with various values of ϵ (0.05, 0.1 and 0.2). All plots are projected on the (x, y) -plane

the same ones in Sec. 3.1. The identified expressions of the quasipotential and circulatory component are

$$U^{\text{Symb}}(\mathbf{x}) = 1.014x^4 - 2.03x^2 + 0.976y^2 + 0.967z^2 + 1.018, \\ \mathbf{g}^{\text{Symb}}(\mathbf{x}) = [-1.017y - 1.024z, 2.232x^3 - 2.232x, 2.229x^3 - 2.23x]^T.$$

In Fig. 5 (b) and (c), we show plots of the predicted trajectories generated using the learned dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$ and the identified potential $U^{\text{Symb}}(\mathbf{x})$. The results demonstrate the capability of the proposed method for dealing with noisy trajectories.

3.1.2 Extension to a 4D system

We extend the three-dimensional non-gradient dynamical system of Eq. (20) to four dimensions by introducing an additional state variable w and its corresponding equation $\dot{w}$, formulated in the same structural form as $\dot{y}$ and $\dot{z}$:

$$\begin{cases} \dot{x} = -2(x^3 - x) - y - z - w, \\ \dot{y} = -y + 2(x^3 - x), \\ \dot{z} = -z + 2(x^3 - x), \\ \dot{w} = -w + 2(x^3 - x), \end{cases} \quad (27)$$

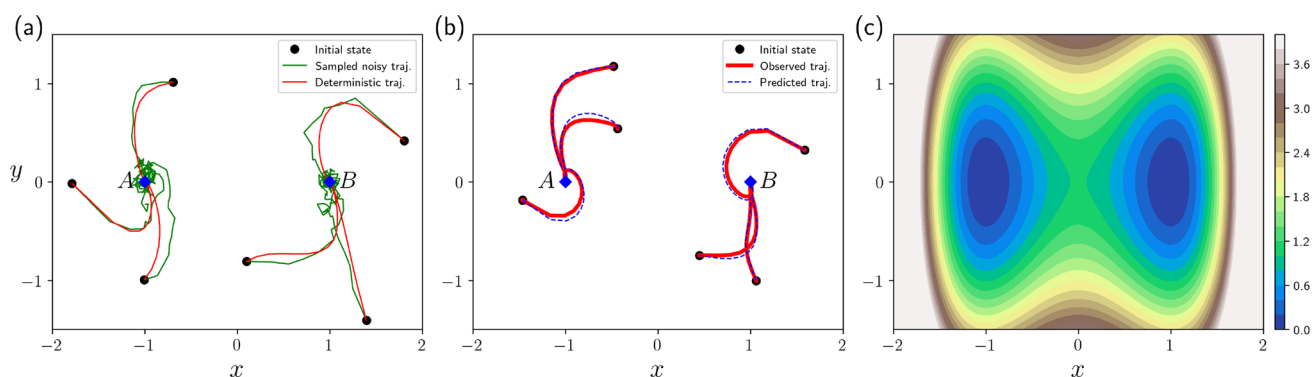


Fig. 5 Combined data-driven method applied to the system (26) with 5000 sampled noisy trajectories. (a) Six generated trajectories and the ones sampled from the deterministic system (20) with the same initial conditions. (b) Six trajectories generated by simulating the learned

dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$ and the ones sampled from the system (20) with the same initial states. (c) The identified quasipotential $U^{\text{Symb}}(x, y, z)$. All plots are projected on the (x, y) -plane

This example follows the same structure as the 3D case in Sec. 3.1, with two stable equilibrium points at $A = (-1, 0, 0, 0)$ and $B = (1, 0, 0, 0)$. The quasipotential consistent with the vector field is given by $U(\mathbf{x}) = x^4 - 2x^2 + y^2 + z^2 + w^2 + 1$. Using the same procedure as in the 3D case, we construct the decomposition into downhill and circulatory components and recover the symbolic form of the quasipotential. A comparison between the data-driven identification and the exact solution is reported in Table 2.

The comparison between the generated trajectories and the analytically reconstructed ones is shown in Fig. 6, illustrating the applicability of the method to quasipotential problems in dimensions higher than three, a topic that has not been thoroughly investigated in the literature.

3.2 Example 2: Nanomechanical graphene resonator

Here, the data-driven technique is applied to the dynamics of a graphene nanomechanical resonators. The equations

$$\begin{cases} \dot{P} = \frac{\omega_0^2 - \omega_F^2}{2\omega_F} Q - \zeta P + \frac{3}{8} \frac{\alpha}{\omega_F} Q (P^2 + Q^2) \\ \dot{Q} = -\frac{\omega_0^2 - \omega_F^2}{2\omega_F} P - \zeta Q - \frac{3}{8} \frac{\alpha}{\omega_F} P (P^2 + Q^2) - \frac{\beta}{2\omega_F} \end{cases} \quad (28)$$

describe the slow dynamics of a graphene membrane in terms of the P and Q variables that are the slowly varying in-phase and out-of-phase components of the motion, and obtained by applying a rotating wave approximation to the actual fast dynamics [6]. We take the following parameters in the dynamical equation: resonant frequency $\omega_0 = 1$, frequency of the excitation $\omega_F = 1.0018$, damping ratio $\zeta = 0.00045$, cubic stiffness coefficient $\alpha = 33$ and amplitude of the excitation $\beta = 1.4 \times 10^{-5}$. These describe a graphene drum in the classical duffing bistable region where both high and low

amplitude oscillations occur. The dynamics has two stable fix points at $A \approx (-0.007, -0.011)$ and $B \approx (0.004, -0.001)$, as shown in Fig. 7 (a).

We create a synthetic dataset $\{X_i^{j,0}, X_i^{j,h}\}$ by generating 2000 trajectories of the system (28) with initial conditions in a domain $\Omega = [-0.02, 0.02] \times [-0.02, 0.02]$. From the data, we learn an orthogonal decomposition of the vector field by training the neural networks $V_\theta(P, Q)$ and $\mathbf{g}_\theta(P, Q)$. Subsequently, we identify symbolic expressions for the components in the decomposition. A target matrix $\mathbf{G}(\mathbf{X})$ is constructed with the neural network solutions over a subset of $\{X_i^{j,0}\}$ built with a potential threshold $\tau = \min_{\Omega} V_\theta(\mathbf{x}) + 6 \times 10^{-9}$. The full dataset used for training the neural networks, along with the reduced dataset utilized during the regression phase, are displayed in Fig. 7 (b) and (c). The basis functions in the identification are taken as polynomials in (P, Q) up to fourth and third orders for V^{Symb} and $\mathbf{g}^{\text{Symb}}$, respectively. We set the parameters as $\lambda = 10^{-9}$ and $\nu = 10^{-2}$ in the problem (17). An initial guess for the coefficient matrix Ξ_v and Ξ_g is obtained by performing regression for $V^{\text{Symb}}(\mathbf{x})$ and $\mathbf{g}^{\text{Symb}}(\mathbf{x})$ separately, as we detail in Appendix A.

The identified symbolic expression (with rounded coefficients) for the quasipotential is

$$\begin{aligned} U^{\text{Symb}}(P, Q) = & (8.12 \times 10^{-9}) \\ & - (1.56 \times 10^{-6}) P + (3.58 \times 10^{-7}) Q \\ & + (9.65 \times 10^{-6}) P^2 + (1.71 \times 10^{-5}) P Q + (5.83 \times 10^{-4}) Q^2 \\ & + 0.023 P^3 + 0.039 P^2 Q + 0.047 P Q^2 + 0.098 Q^3 \\ & + 1.05 P^4 + 1.88 P^3 Q + 3.13 P^2 Q^2 + 2.47 P Q^3 + 4.08 Q^4. \end{aligned} \quad (29)$$

Also, the identified symbolic expression (with rounded coefficients) for the circulatory component, $\mathbf{g}^{\text{Symb}}(P, Q) =$

Table 2 The exact and identified expressions for the quasipotential $U(\mathbf{x})$, downhill component $\nabla V(\mathbf{x})$ and circulatory component $\mathbf{g}(\mathbf{x})$ in the decomposition of the vector field for the system (27). Three digits after the decimal point are displayed in the identified expressions

	Exact expression	Identified expression
$U(\mathbf{x})$	$x^4 - 2x^2 + y^2 + z^2 + w^2 + 1$	$0.976x^4 - 1.972x^2 + 0.951y^2 + 1.007z^2 + 1.014w^2 + 0.994$
$\nabla V(\mathbf{x})$	$\begin{bmatrix} 2x^3 - 2x \\ y \\ z \\ w \end{bmatrix}$	$\begin{bmatrix} 1.952x^3 - 1.972x \\ 0.951y \\ 1.007z \\ 1.014w \end{bmatrix}$
$\mathbf{g}(\mathbf{x})$	$\begin{bmatrix} -y - z - w \\ 2x^3 - 2x \\ 2x^3 - 2x \\ 2x^3 - 2x \end{bmatrix}$	$\begin{bmatrix} -1.0y - 1.014z - 0.993w \\ 1.997x^3 - 1.990x \\ 2.002x^3 - 1.998x \\ 2.016x^3 - 2.015x \end{bmatrix}$

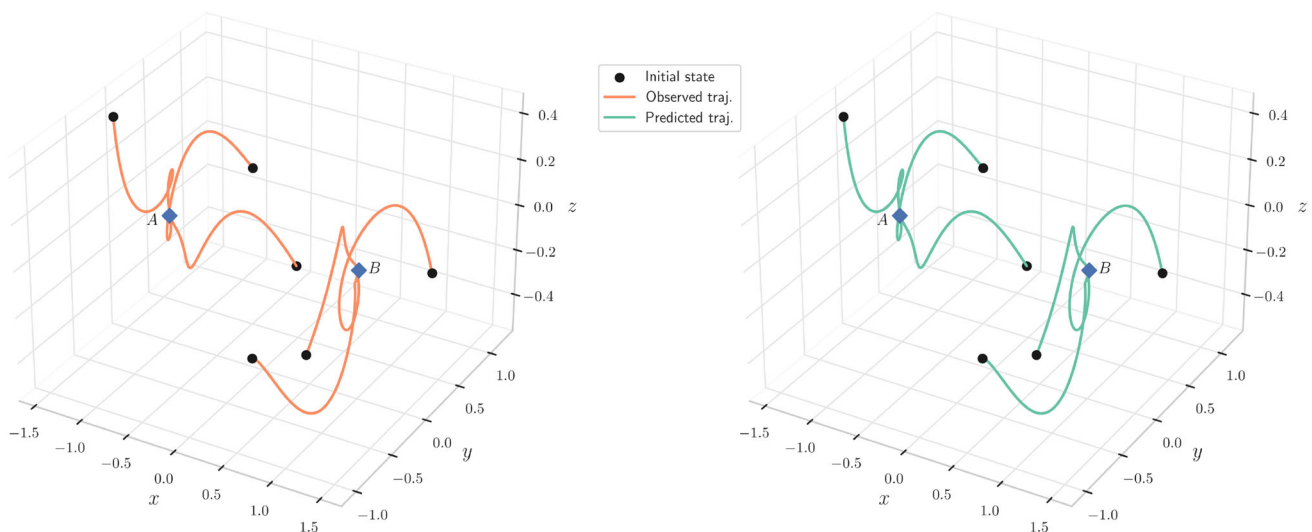


Fig. 6 Phase plots of six observed trajectories (**Left**) and the trajectories generated by simulating the dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$ with the same initial states (**Right**). The plots are projected on the (x, y, z) -space

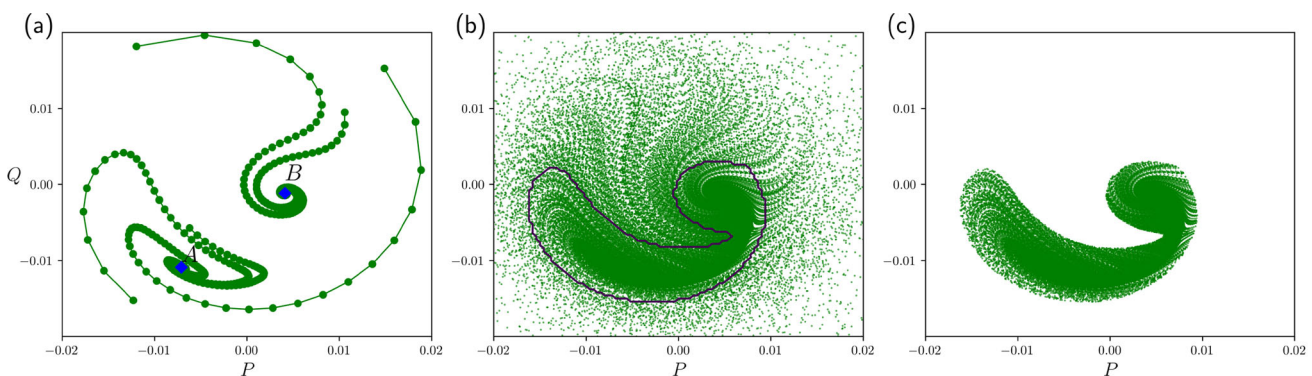


Fig. 7 (a) Five generated trajectories of Eq. (28). (b) The observed dataset $\{X_i^{j,0}\}$, where the solid closed curve indicates the boundary of the region $\{\mathbf{x} \in \Omega : V_\theta(\mathbf{x}) < \tau\}$ with a threshold value $\tau = \min_\Omega V_\theta(\mathbf{x}) + 6 \times 10^{-9}$. (c) The dataset for the data matrix $\mathbf{X}$ sampled using the confined region

$[g_1^{\text{symb}}(P, Q), g_2^{\text{symb}}(P, Q)]$ is

$$g_1^{\text{symb}}(P, Q) = -(5.15 \times 10^{-7}) \\ -(4.58 \times 10^{-4})P - (1.88 \times 10^{-3})Q$$

$$+0.027P^2 + 0.048PQ - 0.005Q^2 \\ +2.01P^3 + 14.61P^2Q + 4.17PQ^2 + 11.99Q^3, \\ g_2^{\text{symb}}(P, Q) = -(6.61 \times 10^{-6}) \\ +(1.77 \times 10^{-3})P - (3.47 \times 10^{-5})Q$$

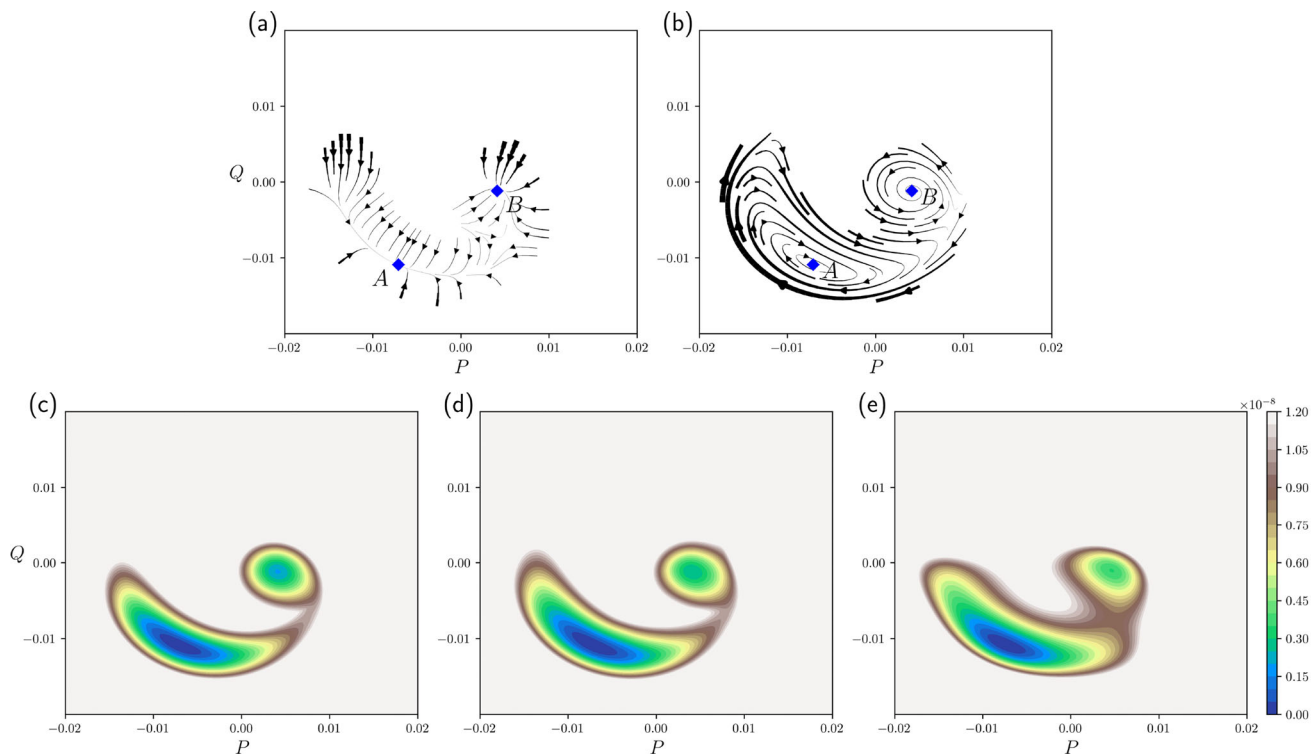


Fig. 8 Combined data-driven method applied to the system (28) with 2000 sampled trajectories. (a) The identified downhill component $-\nabla V^{\text{Symb}}(P, Q)$. (b) The identified circulatory component $\mathbf{g}^{\text{Symb}}(P, Q)$. (c) The quasipotential $U(P, Q)$ estimated using a stan-

dard ordered upwind method [23]. (d) The potential $U_\theta = 2V_\theta(P, Q)$ learned using the neural networks. (e) The potential $U^{\text{Symb}}(P, Q)$ identified using the sparse symbolic regression. In panels (a) and (b), the line thickness shows the flow velocity

$$+0.012P^2 + 0.034PQ + 0.116Q^2 - 11.60P^3 + 2.69P^2Q - 9.68PQ^2 + 6.92Q^3. \quad (30)$$

In Fig. 8 (a) and (b), we show plots of the identified downhill and circulatory components $\nabla V^{\text{Symb}}(P, Q)$, $\mathbf{g}^{\text{Symb}}(P, Q)$ in the decomposition of the vector field. Figure 8 (c) reports a reference solution for the quasipotential U using a standard ordered upwind method with a mesh of 1000×1000 grid points [2]. The mesh-based method is compared with the identification provided by the neural network (Fig. 8 (d)), and the sparse symbolic regression (Fig. 8 (e)). From the comparison with the reference solution, one can observe that the neural network (panel (d)) and symbolic (panel (e)) capture the topology of the energy landscape well. However, we observe a misalignment between the outbound regions of U^{Symb} (Fig. 8 (e)) and the reference solution. We attribute this imprecision to the low data density, as shown in Fig. 7 (b), and the high flow velocity, illustrated in Fig. 8 (a), both of which require special handling during the decomposition step. The error between the reference quasipotential estimated using the standard ordered upwind method (Fig. 5(c)) and the two data-driven steps of our method is presented in Appendix E. This highlights a limitation of the approach: it is fully data-driven and does not

require prior knowledge of the system, but its performance depends heavily on the quantity and quality of the available data. We emphasize that the dataset can be improved through filtering, smoothing, and interpolation techniques to mitigate these issues. Additionally, the representation capacity of the polynomials in the candidate library for symbolic regression may be limited in capturing such variation. Despite this, the solution $U^{\text{Symb}}(P, Q)$ gives a physically interpretable form of the quasipotential. In this example, we also implement the data-driven method in 50 independent runs and obtain similar potential U^{Symb} as in Fig. 8 (e), where plots of the coefficient distribution for each library term in $U^{\text{Symb}}(P, Q)$ are shown in Fig. 10 of Appendix D.

Long-term prediction. Here we analyze the accuracy of the prediction for the identified dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$. Given 1000 observed trajectories of the system (28) with initial states sampled from $\{\mathbf{x} \in \Omega : V_\theta(\mathbf{x}) < \tau\}$, we generate trajectories by evolving the identified dynamics from the states. The statistics of the errors for the generated trajectories, as defined in Eq. (21), indicate that within one standard deviation, the error may be as high as approximately 5%. Fig. 9 (a) shows a comparison of the generated trajectories with the observed ones, demonstrating the capacity to accurately capture the underlying dynamical evolution.

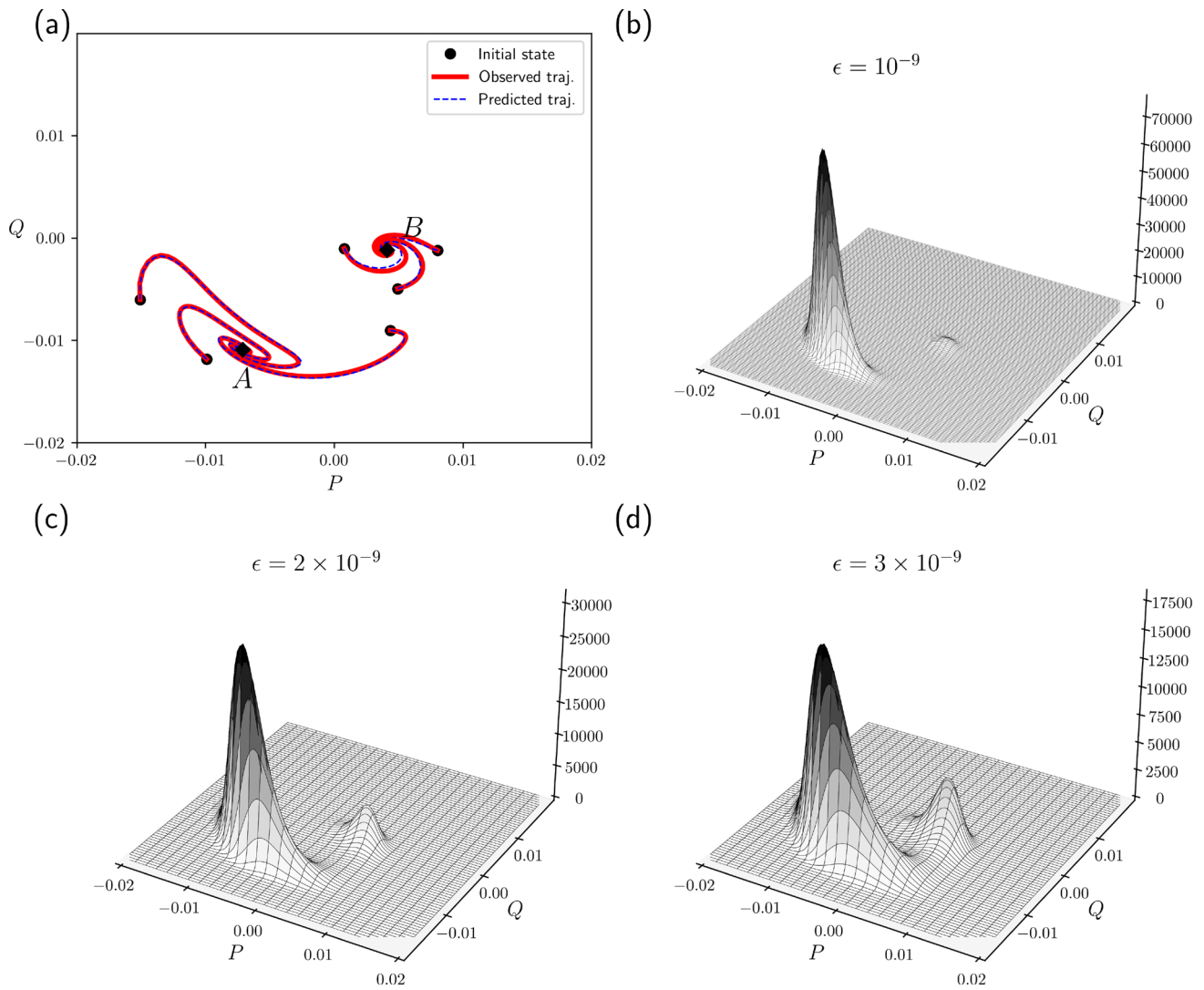


Fig. 9 (a) Phase plot of six observed trajectories and the trajectories generated by simulating the dynamics $\dot{\mathbf{x}} = \mathbf{f}^{\text{Symb}}(\mathbf{x})$ with the same initial states. (b)-(d) Plots of the invariant distribution computed by

$p_\epsilon(P, Q) = Z_\epsilon^{-1} \exp(-U^{\text{Symb}}(P, Q)/\epsilon)$ of the randomly perturbed dynamics system (22) with various values of ϵ (10^{-9} , 2×10^{-9} , and 3×10^{-9})

Invariant distribution. With the identified quasipotential $U^{\text{Symb}}(P, Q)$ of Eq. (29), we compute an explicit invariant distribution for the nanomechanical resonator when perturbed by a random noise. It is important to investigate how fluctuations affect the out-of-equilibrium dynamics of nonlinear resonators because atomically thin resonators are very sensitive to environmental noise. The invariant distribution of the randomly perturbed system is approximated by

$$p_\epsilon(P, Q) = Z_\epsilon^{-1} \exp(-U^{\text{Symb}}(P, Q)/\epsilon). \quad (31)$$

Here the normalization constant is estimated by the integration over the domain Ω . The evolution of the invariant distribution $p_\epsilon(P, Q)$ for different values of ϵ , namely 10^{-9} , 2×10^{-9} and 3×10^{-9} , is showcased in Fig. 9 (b)-(d).

While the predominance of the large oscillations of the nanomechanical resonator persists (motion around the large amplitude attractor - fixed point A), we observe a gradual smearing of the distribution in the phase space of the quadratures for larger ϵ .

4 Conclusions

To conclude, we introduced a computational method for discovering explicit expressions of the quasipotential from observed data, which combines a neural network approximation and sparse regression. The main contribution of this paper is to demonstrate, for the first time, that it is possible to retrieve an analytical form of the quasipo-

tential function using a fully data-driven approach. Two applications are presented. The first example validates the accuracy of the identified quasipotential, while the second highlights the ability of the data-driven method to interpret nanomechanical vibrational systems of practical engineering interest. Unlike existing approaches, our method can explicitly recover quasipotential expressions without any assumptions about the governing vector field. In particular, we do not use any a priori knowledge of the vector field equations. Instead, all information necessary to reconstruct the quasipotential function is derived directly from the data. This renders the method applicable to both synthetic and experimental datasets, as it only requires observations of the nonlinear dynamics. Moreover, the algorithm is robust to noisy data, which we demonstrate through tests mimicking realistic systems subject to perturbations and uncertainties.

The data-driven approach is suitable for a wide range of systems exhibiting complex multistable behavior, provided that sufficient trajectory data are available. Specifically, it requires adequate coverage of the phase space by the input trajectories to ensure accurate reconstruction. Additionally, it demands careful GPU memory management to avoid computational bottlenecks. Because our method relies on neural networks and sparse regression applied directly to available trajectory data rather than a discretized computational grid, its mathematical formulation and implementation fundamentally differ from grid-based methods. This avoids the need to operate on an n -dimensional grid, thereby mitigating the curse of dimensionality and facilitating extension to higher dimensions. We illustrate this capability with a four-dimensional example. Given these advantages, we believe the proposed approach offers a promising tool for precisely predicting key statistical quantities related to transitions between metastable states. This novel idea of determining a symbolic quasipotential function by coupling machine learning techniques presents a fully data-driven solution, demonstrating how recent advances in data science and machine learning open exciting new opportunities in the study of nonlinear dynamics.

Appendix A Initialization routine for the regularized regression

An initial coefficient vector Ξ_v^0 for the potential V is obtained by

$$\arg \min_{\mathbf{W}, \Xi_v} \frac{1}{2} \|V_\theta(\mathbf{X}) - \Theta(\mathbf{X}) \Xi_v\|^2 + \lambda R(\mathbf{W}) + \frac{1}{2\nu} \|\Xi - \mathbf{W}\|^2, \quad (\text{A1})$$

where $V_\theta(\mathbf{X})$ is the target vector for V . Similarly, an initial coefficient matrix Ξ_g^0 for the circulatory component $\mathbf{g}$ is produced by solving

$$\arg \min_{\mathbf{W}, \Xi_g} \frac{1}{2} \|\mathbf{g}_\theta(\mathbf{X}) - \Theta(\mathbf{X}) \Xi_g\|^2 + \lambda R(\mathbf{W}) + \frac{1}{2\nu} \|\Xi - \mathbf{W}\|^2, \quad (\text{A2})$$

where $\mathbf{g}_\theta(\mathbf{X})$ is the target matrix for $\mathbf{g}$. Therefore, with the relation (15) for the vector field $\mathbf{f}$, an initial value for Ξ for the problem (17) is given by

$$\Xi^0 = [-T(\Xi_v^0) + \Xi_g^0, \Xi_v^0, \Xi_g^0]. \quad (\text{A3})$$

Appendix B Data sampling and neural networks training

Data points are collected from N trajectories generated from numerical simulations of the dynamics. We employ a four-order Runge-Kutta integrator with time step $h/5$ and initial states sampled from a uniform distribution over the computational region Ω . Then the dataset $\{X_i^{j,0}, X_i^{j,h}\}$ is constructed by taking snapshots of the generated trajectories at times t_j and $t_j + h$, $1 \leq j \leq M$. In total, there are 5×10^5 points in the dataset for the first example (4×10^5 points for the second example). We build a representative subset $\{\tilde{X}_k\}$ to impose the orthogonality condition as in Eq. (9) using the Algorithm 1 of Ref. [41] with a parameter r of 0.1 for the first example and 5×10^{-4} for the second example.

With the two datasets $\{X_i^{j,0}, X_i^{j,h}\}$ and $\{\tilde{X}_k\}$, we train the neural networks $V_\theta(\mathbf{x})$ and $\mathbf{g}_\theta(\mathbf{x})$ using the loss function of Eq. (9). We utilize fully-connected neural networks for $V_\theta^{\text{NN}}, \mathbf{g}_\theta^{\text{NN}}$ with the activation function being the hyperbolic tangent ($\tanh$). In prior to training the networks, a normalization step and multiplying scalars are added to the input and output of the networks as in Eq. (10) and (11). In the loss function, we set the parameter λ as 10 and 10^{-11} for the first and second examples, respectively, and take the numerical integrator $\mathcal{I}$ as the second order Runge-Kutta scheme. We train the neural networks for 1.5×10^5 steps using the stochastic gradient descent with the Adam optimizer and a batch of 5000 data points. In summary, the parameters used for sampling data and training neural networks are reported in Table 3.

Appendix C Estimation of the normalization constant in the invariant distribution for Example 1

We write the symbolic quasipotential $U^{\text{Symb}}(\mathbf{x})$ in Table 1 as

$$U^{\text{Symb}}(\mathbf{x}) = a_1 x^4 - a_2 x^2 + a_3 + a_4 y^2 + a_5 z^2, \quad (\text{C4})$$

where a_1, a_2, a_3, a_4 and a_5 are constant. The normalization constant in the invariant distribution $p_\epsilon(\mathbf{x}) = Z_\epsilon^{-1} \exp(-U^{\text{Symb}}(\mathbf{x})/\epsilon)$ can be computed by

$$\begin{aligned} Z_\epsilon &= \iiint_{\mathbb{R}^3} \exp(-U^{\text{Symb}}(x, y, z)/\epsilon) dx dy dz \\ &= \int_{\mathbb{R}} \exp\left[-\frac{a_1}{\epsilon} x^4 + \frac{a_2}{\epsilon} x^2 - \frac{a_3}{\epsilon}\right] dx \cdot \int_{\mathbb{R}} \exp \end{aligned}$$

Table 3 The parameters for sampling data and training neural networks in the two numerical examples

	Parameters	Archetypal model	Nanomechanical graphene resonator
Sampling data	Ω	$[-2, 2] \times [-1.5, 1.5] \times [-1.5, 1.5]$	$[-0.02, 0.02] \times [-0.02, 0.02]$
	N	5000	2000
	M	50	100
	h	0.01	10
	t_j	$0.1j$	$100j$
	Numerical integrator	RK4	RK4
V_θ^{NN}	Structure	3-50-50-50-1	2-100-100-100-1
	Activation	tanh	tanh
$\mathbf{g}_\theta^{\text{NN}}$	Structure	3-50-50-50-3	2-100-100-100-2
	Activation	tanh	tanh
Training	Optimizer	Adam	Adam
	Learning rate	Exponentially decays	Exponentially decays
	Batch size	5000	5000

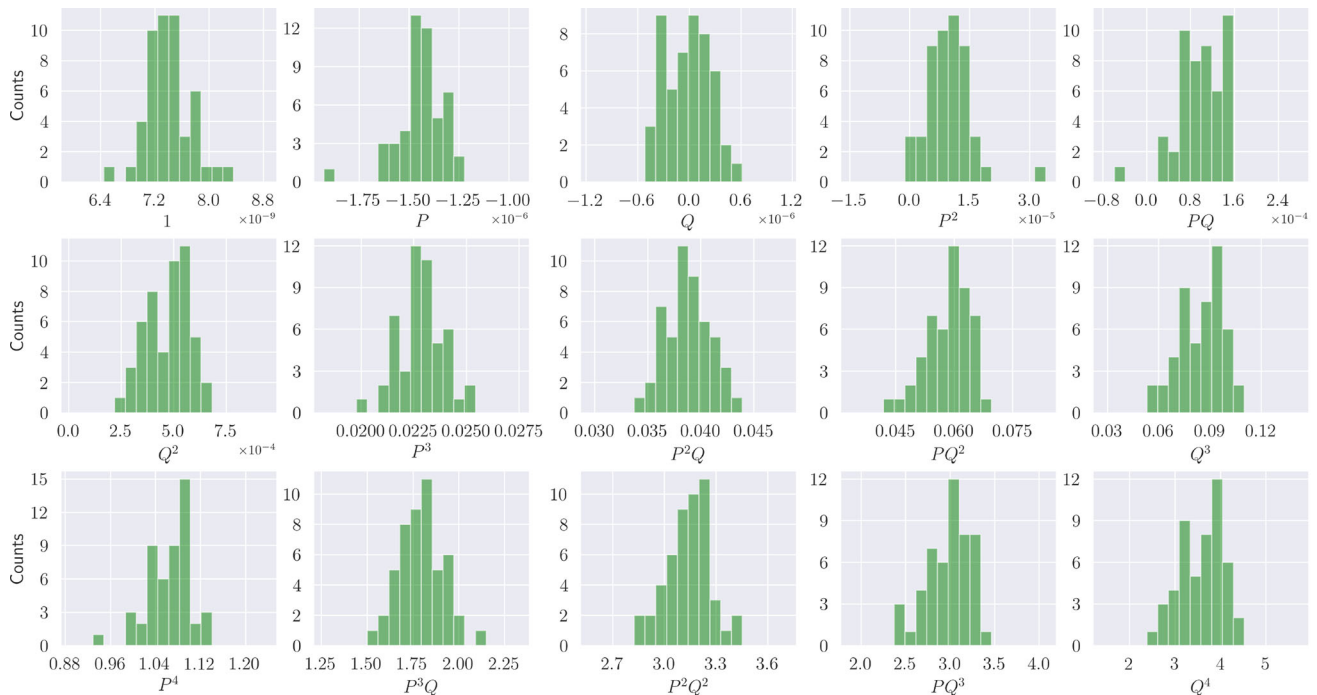
$$\begin{aligned}
& \left[-\frac{a_4}{\epsilon} y^2 \right] dy \cdot \int_{\mathbb{R}} \exp \left[-\frac{a_5}{\epsilon} z^2 \right] dz \\
&= \int_{\mathbb{R}} \exp \left[-\frac{a_1}{\epsilon} x^4 + \frac{a_2}{\epsilon} x^2 - \frac{a_3}{\epsilon} \right] dx \\
& \quad \cdot \sqrt{\pi \epsilon / a_4} \cdot \sqrt{\pi \epsilon / a_5} \\
&= \sqrt{\frac{a_2}{8a_1}} \frac{\pi^2 \epsilon}{\sqrt{a_4 a_5}} \exp \left[\frac{a_2^2}{8a_1 \epsilon} - \frac{a_3}{\epsilon} \right] \\
& \quad \left[I_{1/4} \left(\frac{a_2^2}{8a_1 \epsilon} \right) + I_{-1/4} \left(\frac{a_2^2}{8a_1 \epsilon} \right) \right],
\end{aligned}$$

(C5)

where $I_\alpha(\cdot)$ denotes the modified Bessel function of the first kind [56].

Appendix D Statistics of the learned coefficients for Example 2

Fig. 10 showcases the distributions of the learned coefficients for each library term in $U^{\text{Symb}}(\mathbf{x})$.

**Fig. 10** Distributions of the coefficients for each library term in the identified potential $U^{\text{Symb}}(P, Q)$ in 50 independent runs for of Sec. 3.2

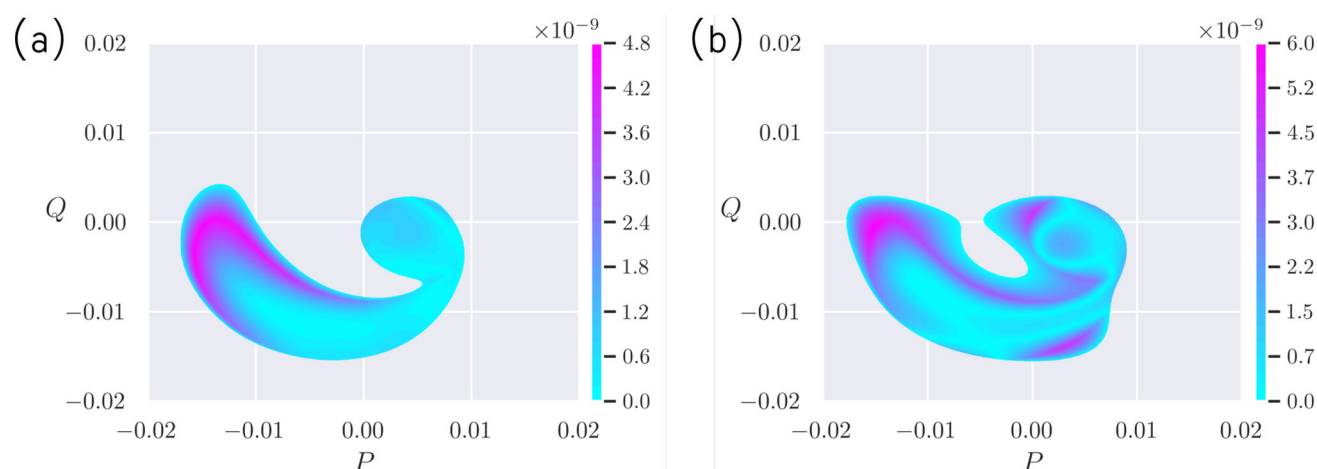


Fig. 11 Error in the estimated data-driven quasipotential. (a) Error of the neural network estimate (Fig. 8 (d)) relative to the ordered upwind method (Fig. 8 (c)). (b) Error of the symbolic regression estimate (Fig. 8 (e)) relative to the ordered upwind method (Fig. 8 (c))

Appendix E Error estimation for Example 2

In Fig. 11 we report the error between the reference quasipotential estimated using the standard ordered upwind method (Fig. 8 (c)) and the two data-driven steps of our method, namely the black-box neural network function, and the analytical expression obtained via symbolic regression.

As shown in Fig. 11, the largest discrepancies occur in the outer regions of the phase space, where the decomposition is challenging to estimate due to the high flow speeds of the underlying vector field.

Author Contributions P.B. and B.L. conceived the research idea. B.L. developed the code and conducted the numerical investigations. P.B. supervised the project. All authors contributed to data analysis, interpretation of the results, and writing of the manuscript.

Funding Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement. The work of BL was partially supported by A*STAR under its AME Programmatic programme: Explainable Physics-based AI for Engineering Modelling & Design (ePAD) [Award No. A20H5b0142].

PB acknowledges partial support from the European Union's NextGenerationEU programme, in the framework of PRIN 2022, project DIMIN.

Data Availability All computational results are reproducible and code can be found at <https://github.com/LinBoNUS/SIQ>

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material

is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Beisner, B., Haydon, D., Cuddington, K.: Alternative stable states in ecology. *Front. Ecol. Environ.* **1**(7), 376–382 (2003). [https://doi.org/10.1890/1540-9295\(2003\)001\[0376:ASSIE\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2003)001[0376:ASSIE]2.0.CO;2)
2. Nolting, B.C., Abbott, K.C.: Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology* **97**(4), 850–864 (2016). <https://doi.org/10.1890/15-1047.1>
3. Wales, D.: *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge Molecular Science. Cambridge University Press, Cambridge (2003)
4. Wang, J., Xu, L., Wang, E.: Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci.* **105**(34), 12271–12276 (2008). <https://doi.org/10.1073/pnas.0800579105>
5. Chan, H.B., Dykman, M.I., Stambaugh, C.: Paths of fluctuation induced switching. *Phys. Rev. Lett.* **100**, 130602 (2008). <https://doi.org/10.1103/PhysRevLett.100.130602>
6. Dolleman, R.J., Belardinelli, P., Houri, S., van der Zant, H.S.J., Alijani, F., Steeneken, P.G.: High-Frequency Stochastic Switching of Graphene Resonators Near Room Temperature. *Nano Lett.* **19**(2), 1282–1288 (2019). <https://doi.org/10.1021/acs.nanolett.8b04862>
7. Huber, J.S., Rastelli, G., Seitner, M.J., Kölbl, J., Belzig, W., Dykman, M.I., Weig, E.M.: Spectral Evidence of Squeezing of a Weakly Damped Driven Nanomechanical Mode. *Phys. Rev. X* **10**(2), 021066 (2020). <https://doi.org/10.1103/PhysRevX.10.021066>
8. McCann, L.I., Dykman, M., Golding, B.: Thermally activated transitions in a bistable three-dimensional optical trap. *Nature* **402**(6763), 785–787 (1999). <https://doi.org/10.1038/45492>
9. Hanggi, P.: Escape from a metastable state. *J. Stat. Phys.* **42**(1), 105–148 (1986). <https://doi.org/10.1007/BF01010843>
10. Militaru, A., Innerbichler, M., Frimmer, M., Tebbenjohanns, F., Novotny, L., Dellago, C.: Escape dynamics of active particles in multistable potentials. *Nat. Commun.* **12**(1), 2446 (2021). <https://doi.org/10.1038/s41467-021-22647-6>
11. Dumont, V., Bestler, M., Catalini, L., Margiani, G., Zilberberg, O., Eichler, A.: Energy landscape and flow dynamics measurements

- of driven-dissipative systems. *Phys. Rev. Res.* **6**, 043012 (2024). <https://doi.org/10.1103/PhysRevResearch.6.043012>
12. Mello, C.C., Barrick, D.: An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci.* **101**(39), 14102–14107 (2004). <https://doi.org/10.1073/pnas.0403386101>
 13. Peters, M., Zhao, T., George, S., Truong, V.G., Nic Chormaic, S., Ying, C., Nome, R., Gordon, R.: Energy landscape of conformational changes for a single unmodified protein. *npj Biosensing* **1**(1), 14 (2024) <https://doi.org/10.1038/s44328-024-00014-x>
 14. Fan, Y., Iwashita, T., Egami, T.: Energy landscape-driven non-equilibrium evolution of inherent structure in disordered material. *Nat. Commun.* **8**(1), 15417 (2017). <https://doi.org/10.1038/ncomms15417>
 15. Ashwin, P., Wieczorek, S., Vitolo, R., Cox, P.: Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **370**(1962), 1166–1184 (2012). <https://doi.org/10.1098/rsta.2011.0306>
 16. McKay, D.I.A., Staal, A., Abrams, J.F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S.E., Rockström, J., Lenton, T.M.: Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science* **377**(6611), 7950 (2022). <https://doi.org/10.1126/science.abn7950>
 17. Chapman, R.R., Ashwin, P., Wood, R.A.: Quantifying tipping behavior: Geometric early warnings and quasipotentials for a box model of amoc. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **35**(2), 023164 (2025). <https://doi.org/10.1063/5.0240959>
 18. Schön, J.C.: Energy landscapes—past, present, and future: A perspective. *J. Chem. Phys.* **161**(5), 050901 (2024). <https://doi.org/10.1063/5.0212867>
 19. Suzuki, K., Nakaoka, S., Fukuda, S., Masuya, H.: Energy landscape analysis elucidates the multistability of ecological communities across environmental gradients. *Ecol. Monogr.* **91**(3), 01469 (2021). <https://doi.org/10.1002/ecm.1469>
 20. Sánchez-Pinillos, M., Dakos, V., Kéfi, S.: Ecological dynamic regimes: A key concept for assessing ecological resilience. *Biological Conservation* **289**, 110409 (2024). <https://doi.org/10.1016/j.biocon.2023.110409>
 21. Manuel, A.P., Lambert, J., Woodside, M.T.: Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. *Proc. Natl. Acad. Sci.* **112**(23), 7183–7188 (2015). <https://doi.org/10.1073/pnas.1419490112>
 22. Gupta, A.N., Vincent, A., Neupane, K., Yu, H., Wang, F., Woodside, M.T.: Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat. Phys.* **7**(8), 631–634 (2011). <https://doi.org/10.1038/nphys2022>
 23. Cameron, M.K.: Finding the quasipotential for nongradient sdes. *Physica D* **241**(18), 1532–1550 (2012). <https://doi.org/10.1016/j.physd.2012.06.005>
 24. Freidlin, M.I., Wentzell, A.D.: *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften, vol. v. 260. Springer, New York (1998)
 25. Dykman, M.I., Maloney, C.M., Smelyanskiy, V.N., Silverstein, M.: Fluctuational phase-flip transitions in parametrically driven oscillators. *Phys. Rev. E* **57**, 5202–5212 (1998). <https://doi.org/10.1103/PhysRevE.57.5202>
 26. Kim, W.K., Netz, R.R.: The mean shape of transition and first-passage paths. *J. Chem. Phys.* **143**(22), 224108 (2015). <https://doi.org/10.1063/1.4936408>
 27. Bouchet, F., Reygner, J.: Generalisation of the eyring-kramers transition rate formula to irreversible diffusion processes. *Ann. Henri Poincaré* **17**(12), 3499–3532 (2016). <https://doi.org/10.1007/s00023-016-0507-4>
 28. Sethian, J.A., Vladimirov, A.: Ordered upwind methods for static hamilton-jacobi equations. *Proc. Natl. Acad. Sci.* **98**(20), 11069–11074 (2001). <https://doi.org/10.1073/pnas.201222998>
 29. Sethian, J.A., Vladimirov, A.: Ordered upwind methods for static hamilton-jacobi equations: Theory and algorithms. *SIAM J. Numer. Anal.* **41**(1), 325–363 (2004)
 30. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. *Proc. Natl. Acad. Sci.* **93**(4), 1591–1595 (1996). <https://doi.org/10.1073/pnas.93.4.1591>
 31. Potter, S.F., Cameron, M.K.: *Ordered Line Integral Methods for Solving the Eikonal Equation* (2019)
 32. Dahiya, D., Cameron, M.: Ordered line integral methods for computing the quasi-potential. *J. Sci. Comput.* **75**(3), 1351–1384 (2018). <https://doi.org/10.1007/s10915-017-0590-9>
 33. Yang, S., Potter, S.F., Cameron, M.K.: Computing the quasipotential for nongradient sdes in 3d. *Journal of Computational Physics* **379**, 325–350 (2019). <https://doi.org/10.1016/j.jcp.2018.12.005>
 34. Weinan, E., Ren, W., Vanden-Eijnden, E.: Minimum action method for the study of rare events. *Communications on Pure and Applied Mathematics* **57**(5), 637–656 (2004). <https://doi.org/10.1002/cpa.20005>
 35. Zhou, X., Ren, W., E, W.: Adaptive minimum action method for the study of rare events. *The Journal of Chemical Physics* **128**(10), 104111 (2008) <https://doi.org/10.1063/1.2830717>
 36. Kikuchi, L., Singh, R., Cates, M.E., Adhikari, R.: Ritz method for transition paths and quasipotentials of rare diffusive events. *Phys. Rev. Res.* **2**, 033208 (2020). <https://doi.org/10.1103/PhysRevResearch.2.033208>
 37. Heymann, M., Vanden-Eijnden, E.: The geometric minimum action method: A least action principle on the space of curves. *Commun. Pure Appl. Math.* **61**(8), 1052–1117 (2008). <https://doi.org/10.1002/cpa.20238>
 38. Tang, Y., Kurths, J., Lin, W., Ott, E., Kocarev, L.: Introduction to Focus Issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30**(6), 063151 (2020). <https://doi.org/10.1063/5.0016505>
 39. Szalai, R.: Data-driven reduced order models using invariant foliations, manifolds and autoencoders. *Journal of Nonlinear Science* **33**(5), 75 (2023). <https://doi.org/10.1007/s00332-023-09932-y>
 40. Chandrashekar, A., Belardinelli, P., Bessa, M.A., Staufer, U., Alijani, F.: Quantifying nanoscale forces using machine learning in dynamic atomic force microscopy. *Nanoscale Adv.* **4**, 2134–2143 (2022). <https://doi.org/10.1039/D2NA00011C>
 41. Lin, B., Li, Q., Ren, W.: A data driven method for computing quasipotentials. *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*. PMLR, 145, 652–670 (2022)
 42. Lin, B., Li, Q., Ren, W.: Computing the invariant distribution of randomly perturbed dynamical systems using deep learning. *J. Sci. Comput.* **91**(3), 77 (2022). <https://doi.org/10.1007/s10915-022-01844-5>
 43. Lin, B., Li, Q., Ren, W.: Computing high-dimensional invariant distributions from noisy data. *J. Comput. Phys.* **474**, 111783 (2023). <https://doi.org/10.1016/j.jcp.2022.111783>
 44. Li, Y., Xu, S., Duan, J., Liu, X., Chu, Y.: A machine learning method for computing quasi-potential of stochastic dynamical systems. *Nonlinear Dyn.* **109**(3), 1877–1886 (2022). <https://doi.org/10.1007/s11071-022-07536-x>
 45. Li, Y., Yuan, S., Xu, S.: Controlling mean exit time of stochastic dynamical systems based on quasipotential and machine learning. *Communications in Nonlinear Science and Numerical Simulation* **126**, 107425 (2023). <https://doi.org/10.1016/j.cnsns.2023.107425>
 46. Kaptanoglu, A.A., Silva, B.M., Fasel, U., Kaheman, K., Goldschmidt, A.J., Callahan, J.L., Delahunt, C.B., Nicolaou, Z.G., Champion, K., Loiseau, J.-C., et al.: Pysindy: A comprehensive

- python package for robust sparse system identification. arXiv preprint [arXiv:2111.08481](https://arxiv.org/abs/2111.08481) (2021)
47. Champion, K., Zheng, P., Aravkin, A.Y., Brunton, S.L., Kutz, J.N.: A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access* **8**, 169259–169271 (2020)
 48. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **113**(15), 3932–3937 (2016). <https://doi.org/10.1073/pnas.1517384113>
 49. Brackston, R.D., Wynn, A., Stumpf, M.P.H.: Construction of quasipotentials for stochastic dynamical systems: An optimization approach. *Phys. Rev. E* **98**, 022136 (2018). <https://doi.org/10.1103/PhysRevE.98.022136>
 50. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016). <http://www.deeplearningbook.org>
 51. Alpaydin, E.: *Introduction to Machine Learning*. MIT Press (2020)
 52. Quade, M., Abel, M., Nathan Kutz, J., Brunton, S.L.: Sparse identification of nonlinear dynamics for rapid model recovery. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**(6), 063116 (2018). <https://doi.org/10.1063/1.5027470>
 53. Zheng, P., Askham, T., Brunton, S.L., Kutz, J.N., Aravkin, A.Y.: A unified framework for sparse relaxed regularized regression: Sr3. *IEEE Access* **7**, 1404–1423 (2019). <https://doi.org/10.1109/ACCESS.2018.2886528>
 54. Lathourakis, C., Cicirello, A.: Physics enhanced sparse identification of dynamical systems with discontinuous nonlinearities. *Nonlinear Dyn.* **112**(13), 11237–11264 (2024). <https://doi.org/10.1007/s11071-024-09652-2>
 55. Mangan, N.M., Askham, T., Brunton, S.L., Kutz, J.N., Proctor, J.L.: Model selection for hybrid dynamical systems via sparse regression. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475**(2223), 20180534 (2019). <https://doi.org/10.1098/rspa.2018.0534>
 56. Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* 55, (1968)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.