Identification of moment equations via data-driven approaches in nonlinear Schrödinger models

Su Yang[1]*, Shaoxuan Chen[1], Wei Zhu[1,2] and P. G. Kevrekidis[1]

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA, United States, [2]School of Mathematics, Georgia Institute of Technology, Atlanta, GA, United States

Introduction: The moment quantities associated with the nonlinear Schrödinger equation offer important insights into the evolution dynamics of such dispersive wave partial differential equation (PDE) models. The effective dynamics of the moment quantities are amenable to both analytical and numerical treatments.

Methods: In this paper, we present a data-driven approach associated with the "Sparse Identification of Nonlinear Dynamics" (SINDy) to capture the evolution behaviors of such moment quantities numerically.

Results and Discussion: Our method is applied first to some well-known closed systems of ordinary differential equations (ODEs) which describe the evolution dynamics of relevant moment quantities. Our examples are, progressively, of increasing complexity and our findings explore different choices within the SINDy library. We also consider the potential discovery of coordinate transformations that lead to moment system closure. Finally, we extend considerations to settings where a closed analytical form of the moment dynamics is not available.

KEYWORDS

NLS models, SINDy, data-driven methods, moment equations, reduced-order modeling

1 Introduction and motivation

The study of Nonlinear Schrödinger (NLS) type models (Sulem and Sulem, 1999; Ablowitz et al., 2004) is of wide interest and significance in a diverse array of physical modeling settings (Ablowitz and Clarkson, 1991; Ablowitz, 2011). Relevant areas of application extend from atomic physics (Pitaevskii and Stringari, 2003; Pethick and Smith, 2002; Kevrekidis et al., 2015) to fluid mechanical and hydrodynamic (notably ones stemming from deep water waves) problems (Ablowitz, 2011; Infeld and Rowlands, 2000) and from plasma physics (Kono and Skorić, 2010; Infeld and Rowlands, 2000) to nonlinear optics (Kivshar and Agrawal, 2003; Hasegawa and Kodama, 1995). Indeed, the relevant model is a prototypical envelope wave equation that describes the dynamics of dispersive waves. Specifically, in the context of nonlinear optics, it describes the envelope of the electric field of light in optical fibers (as well as waveguides), with the relevant measurable quantity being the light intensity I proportional to the square modulus of the complex field $u(x,t)$. Generalizations of relevant optical applications, involving multiple polarizations or frequencies of light have also been widely considered, in both spatially homogeneous and spatially heterogeneous media (Kivshar and Agrawal, 2003; Hasegawa and Kodama, 1995).

It is worthwhile to note that even to this day, the subject of NLS-type models continues to fascinate researchers and to constitute a fertile platform for a wide range of physical, mathematical and computational works; see, e.g., Karjanto (2024) for a recent review. At the same time, over the past few years, there has been an explosion of interest in data-driven methods, whereby machine-learning techniques are brought to bear towards understanding, codifying and deducing the fundamental quantities of physical systems. Arguably, a turning point in this effort was the development of the so-called physics-informed neural networks (PINNs) by Raissi et al. (2019) and of similar methodologies such as the extension of PINNs via the so-called DeepXDE Lu et al. (2021), as well as the parallel track of sparse identification of nonlinear systems, so-called SINDy by Brunton et al. (2016), which is central to the considerations herein. Additional methods include, but are not limited to, the sparse optimization of Schaeffer (2017), meta-learning of Feliu-Faba et al. (2020), as well as the neural operators of Li et al. (2021). A review of relevant model identification techniques can be found, e.g., in Karniadakis et al. (2021). Notice that a parallel track to the above one seeks not to discover the models, but rather key features thereof such as conserved quantities (Liu and Tegmark, 2021; 2022; Liu et al., 2022; Zhu et al., 2023) and its potential integrability (Krippendorf et al., 2021; De Koster and Wahls, 2024).

In the present setting, we seek to combine this important class of dispersive wave models within optical (and other physical) applications with some of the above machine-learning toolboxes. Our aim is not to discover the full PDE, or its conservation laws/integrability as in some of the above works. Rather, our aim is to leverage the theoretical understanding that exists at the level of reduced-order ODEs in the form of moment methods (Pérez-García et al., 2007; García-Ripoll and Pérez-García, 1999). Indeed, it is well-known from these works that upon defining suitable moment quantities, one can obtain closed form ODE dynamical systems of a few degrees-of-freedom, often just two (lending themselves to dynamical systems analysis) or sometimes involving a few more but still offering valuable low-dimensional analytical insights on the evolution of the center of mass, variance, kurtosis etc. of the relevant distribution. It is those effective ODEs (that stem from the original PDE via the moments) that we aim to retrieve using the data-driven approach developed herein. It is also relevant to mention in passing here that the moment methods were also used successfully to other models such as Fisher-KPP equations, considering also applications to the dynamics, e.g., of brain tumors (Belmonte-Beitia et al., 2014). Our approach and presentation will be structured hereafter as follows. In section 2, we will present a "refresher" from a theoretical perspective of the method of moments, essentially revisiting some basic results from the work of Pérez-García et al. (2007); García-Ripoll and Pérez-García (1999). Then, in Section 3, we will give a brief overview of SINDy type methods and the types of choices (such as, e.g., of model libraries) that they necessitate. Additionally, we will introduce a data-driven approach for learning coordinate transformations to close moment systems when the initially chosen moments are not closed. Then, in Section 4, we present a palette of numerical results and their effective moment identification. Our narrative contains a gradation of examples from simpler ones (where, e.g., an analytical low-dimensional closure of moments may exist) to

gradually more complex ones, where a closure may exist after a coordinate transformation and eventually to cases where a closed system does not exist at the moment level to the best of our knowledge. Our aim is to showcase not only the successes, but also the challenges that the method may encounter in cases where we do not know of a closure or when we may not rightfully choose the library of functions (even when a closure may exist). We hope that this will provide a more informed/balanced viewpoint to the reader about what these methods may (and what they may not) be expected to provide.

2 Moment equation theoretical background

To contextualize our perspective, we will focus on the following specific case of the $(1 + 1)$-dimensional nonlinear Schrödinger (NLS) equation with a harmonic potential, $V(x,t) = \frac{1}{2}x^2$ (Kevrekidis et al., 2015; Kivshar and Agrawal, 2003),

$$iu_t = -\frac{1}{2}u_{xx} + \frac{1}{2}x^2 u + g(|u|^2, t)u, \qquad (1)$$

where $g(|u|^2, t)$ denotes the nonlinearity. This model is not only of relevance to optics (where the harmonic potential represents the heterogeneous profile of the refractive index) (Kivshar and Agrawal, 2003), but also to atomic Bose-Einstein condensates, where this parabolic confinement is a typical byproduct of magnetic traps (Pitaevskii and Stringari, 2003; Kevrekidis et al., 2015). We will consider the initial value problem of Equation 1 with localized and sufficiently regular initial conditions (ICs) $u_0(x) = u(x, 0)$.

2.1 The method of moments

Instead of fully characterizing the solution of the Cauchy problem of Equation 1, the *method of moments* (Pérez-García et al., 2007), seeks to provide qualitative description of the solution behavior by studying the evolution of several integral quantities, i.e., the moments, of the solution $u(x, t)$. This approach enables a reduced-order description of the NLS equation by transforming it into a system of (potentially) closed ordinary differential equations (ODEs). More specifically, according to Pérez-García et al. (2007), we define for $k = 0, 1, 2, \cdots$ the *moment* quantities of solution $u(x, t)$ as follows,

$$I_k(t) = \int_{\mathbb{R}} x^k |u(x,t)|^2 dx, \qquad (2)$$

$$V_k(t) = 2^{k-1} i \int_{\mathbb{R}} x^k \left(u(x,t) \frac{\partial \bar{u}(x,t)}{\partial x} - \bar{u}(x,t) \frac{\partial u(x,t)}{\partial x} \right) dx, \quad (3)$$

$$K(t) = \frac{1}{2} \int_{\mathbb{R}} \left| \frac{\partial u(x,t)}{\partial x} \right|^2 dx, \qquad (4)$$

$$J(t) = \int_{\mathbb{R}} G(\rho(x,t), t) dx = \int_{\mathbb{R}} G(|u(x,t)|^2, t) dx, \qquad (5)$$

where $\bar{u}$ is the complex conjugate of u, $\rho(x,t) = |u(x,t)|^2$, and $G = G(\rho, t)$ is a function such that $\frac{\partial G}{\partial \rho}(\rho, t) = g(\rho, t)$. Moments Equations 2–5 of the solution $u(x, t)$ have intuitive physical meanings; for example, the first moment $I_1(t)$ is associated with the center of mass as described by the (unnormalized) probability

density $\rho = |u|^2$. Higher moments I_k are also associated with this density distribution (i.e., its variance etc.). The V_k quantities are the respective ones associated with the momentum density (which is the quantity in the corresponding parenthesis in the right hand side of the V_k definition. K stems from (thinking quantum-mechanically) the kinetic part of the Schrödinger problem energy, while J represents the nonlinear part of the corresponding energy. We assume that the IC $u_0(x)$ is regular enough to ensure that all moments are well-defined for $t \geq 0$.

The method of moments aims to extract qualitative information about the solution $u(x,t)$ of the PDE Equation 1 by deriving a *closed* set of evolution ODEs for the moments of $u(x,t)$. Depending on the nonlinearity $g(\rho,t)$, these ODEs can sometimes be determined *analytically* as is shown in the work of Pérez-García et al. (2007); García-Ripoll and Pérez-García (1999). Below, we provide a few examples.

Example 1. The moments I_1 and V_0 satisfy

$$\begin{cases} \dfrac{dI_1}{dt} = V_0, \\[2mm] \dfrac{dV_0}{dt} = -I_1. \end{cases} \qquad (6)$$

This indicates that the evolution of the center of mass I_1 behaves as a harmonic oscillator, independent of the nonlinearity $g(\rho,t)$. More generally, for a parabolic confinement of frequency Ω, this would be reflected in the associated frequency of moment oscillations; this is the so-called dipolar motion (Pitaevskii and Stringari, 2003; Kevrekidis et al., 2015).

Example 2. If $g(\rho,t) \equiv 0$, i.e., for the linear case of Equation 1, the set of moments I_2, V_1, K are closed under

$$\begin{cases} \dfrac{dI_2}{dt} = V_1, \\[2mm] \dfrac{dV_1}{dt} = 4K - 2I_2, \\[2mm] \dfrac{dK}{dt} = -\dfrac{1}{2}V_1. \end{cases}$$

Example 3. Assume the nonlinearity $g(\rho,t) = g(\rho)$ is time-independent and given by $g(\rho) = g_0\rho^2$, where $g_0 \in \mathbb{R}$ is a constant. Although the evolution of the moments I_2, V_1, K and J is not closed, it becomes closed under the coordinate transformation $E = K + J$, i.e.,

$$\begin{cases} \dfrac{dI_2}{dt} = V_1, \\[2mm] \dfrac{dV_1}{dt} = 4E - 2I_2, \\[2mm] \dfrac{dE}{dt} = -\dfrac{1}{2}V_1, \end{cases} \qquad (7)$$

While the examples and conditions under which the method of moments leads to closed equations are well-known, deriving such analytical closure systems requires knowledge of the underlying PDE system, Equation 1 and detailed calculations therewith. This work explores *data-driven* methods for obtaining analytical or

approximate moment closure systems based on empirical *observations* or *simulations* of the NLS equation, rather than relying solely on such analytical understanding and derivations. Importantly, the reconstruction of these ODE models can, in principle, take place even for settings where the underlying PDE model is unavailable/has not been specified. Given "experimental" data for the field, one may aspire to utilize the toolboxes presented below in order to obtain these effective, reduced dynamical equations.

For systems with existing analytical closures of the moment equations, such as Examples 1–3, our method seeks to rediscover the moment evolution equations and potentially the necessary coordinate transformations, such as $E = K + J$ in Example 3, in a model-agnostic and data-driven manner. For systems lacking analytical closed moment equations, we seek to derive approximate moment closure equations, providing a principled and reduced-order description of the original PDE, capable of predicting the future evolution of the system. The relevant details will be explained in the following section.

3 Data-driven methods

We present two computational methodologies for finding analytical or approximate closures for moment equations.

3.1 Sparse Identification of Nonlinear Dynamics

Our first method leverages Sparse Identification of Nonlinear Dynamics (SINDy) by Brunton et al. (2016), a data-driven approach for discovering governing ODEs from simulated or observational data. Consider a nonlinear ODE system of the form:

$$\frac{d\mathbf{x}}{dt}(t) = \mathbf{f}(\mathbf{x}(t)), \qquad (8)$$

where $\mathbf{x} = (x_1, \ldots, x_n)^\top : [0, \infty) \to \mathbb{R}^n$ represents the state evolution over time, governed by the dynamic constraint $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$. SINDy aims to identify the unknown dynamics, $\mathbf{f}(\mathbf{x})$ from a time series of $\mathbf{x}$.

The key assumption is that $\mathbf{f}(\mathbf{x})$ has a "simple" form and can be expressed or approximated as a linear combination of only a few terms from a suitably chosen library, $\boldsymbol{\Theta}(\mathbf{x}) = [\theta_1(\mathbf{x}), \ldots, \theta_p(\mathbf{x})]$. For example, a monomial library of degree up to two, $\boldsymbol{\Theta}_{\deg \leq 2}(\mathbf{x})$, is:

$$\begin{aligned} \boldsymbol{\Theta}_{\deg \leq 2}(\mathbf{x}) &= \left[\boldsymbol{\Theta}_{\deg=1}(\mathbf{x}), \boldsymbol{\Theta}_{\deg=2}(\mathbf{x})\right] \\ &= \Big[\underbrace{x_1, x_2, \ldots, x_n}_{\boldsymbol{\Theta}_{\deg=1}(\mathbf{x})}, \underbrace{x_1^2, x_1 x_2, \ldots, x_1 x_n, x_2^2, \ldots, x_n^2}_{\boldsymbol{\Theta}_{\deg=2}(\mathbf{x})}\Big] \\ &= \left[\theta_1(\mathbf{x}), \ldots, \theta_p(\mathbf{x})\right], \qquad (9) \end{aligned}$$

with $p = n + \binom{n}{2}$. In particular, e.g., the right-hand sides of Equations 6, 7 can all be written as sparse linear combinations of terms from $\boldsymbol{\Theta}_{\deg \leq 2}(\mathbf{x})$. Given a dictionary $\boldsymbol{\Theta}(\mathbf{x}) = [\theta_1(\mathbf{x}), \ldots, \theta_p(\mathbf{x})]$ of p elements—p is typically larger than n—the sparsity assumption implies the existence of a sparse matrix $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n) \in \mathbb{R}^{p \times n}$ such that

$$\mathbf{f}(\mathbf{x})^{\top} = \big[\, f_1(\mathbf{x}), \cdots, \; f_n(\mathbf{x}) \,\big] \approx \mathbf{\Theta}(\mathbf{x}) \cdot \mathbf{\Xi}$$

$$= \big[\theta_1(\mathbf{x}), \ldots, \theta_p(\mathbf{x})\big] \cdot \begin{bmatrix} | & | & \cdots & | \\ \xi_1 & \xi_2 & \cdots & \xi_n \\ | & | & \cdots & | \end{bmatrix} \qquad (10)$$

where each sparse column $\xi_j \in \mathbb{R}^p$ indicates which nonlinear functions among the library $\mathbf{\Theta}(\mathbf{x}) = [\theta_1(\mathbf{x}), \ldots, \theta_p(\mathbf{x})]$ are used to parsimoniously represent $f_j(\mathbf{x})$.

To determine this sparse $\mathbf{\Xi}$, SINDy employs sparse regressions on the data. Specifically, given a time series $\{\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N)\} \subset \mathbb{R}^n$ of the state $\mathbf{x}(t)$ at times $t_1, \ldots, t_N$—N is generally much larger than p, the library size—one can assemble the *state matrix* $\mathbf{X} \in \mathbb{R}^{N \times n}$ and the *derivative matrix* $\dot{\mathbf{X}} \in \mathbb{R}^{N \times n}$:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \\ | & | & & | \end{bmatrix} := \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_N) & x_2(t_N) & \cdots & x_n(t_N) \end{bmatrix} \in \mathbb{R}^{N \times n},$$

$$\frac{d}{dt}\mathbf{X} = \dot{\mathbf{X}}$$

$$= \begin{bmatrix} | & | & & | \\ \dot{\mathbf{X}}_1 & \dot{\mathbf{X}}_2 & \cdots & \dot{\mathbf{X}}_n \\ | & | & & | \end{bmatrix} := \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_N) & \dot{x}_2(t_N) & \cdots & \dot{x}_n(t_N) \end{bmatrix} \in \mathbb{R}^{N \times n},$$

where $\dot{\mathbf{X}}$ can be estimated by, e.g., finite differences on $\mathbf{X}$. Define the *library matrix* $\mathbf{\Theta}(\mathbf{X}) \in \mathbb{R}^{N \times p}$ as

$$\mathbf{\Theta}(\mathbf{X}) = \begin{bmatrix} | & | & & | \\ \theta_1(\mathbf{X}) & \theta_2(\mathbf{X}) & \cdots & \theta_p(\mathbf{X}) \\ | & | & & | \end{bmatrix}$$

$$:= \begin{bmatrix} \theta_1(\mathbf{x}(t_1)) & \theta_2(\mathbf{x}(t_1)) & \cdots & \theta_p(\mathbf{x}(t_1)) \\ \theta_1(\mathbf{x}(t_2)) & \theta_2(\mathbf{x}(t_2)) & \cdots & \theta_p(\mathbf{x}(t_2)) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1(\mathbf{x}(t_N)) & \theta_2(\mathbf{x}(t_N)) & \cdots & \theta_p(\mathbf{x}(t_N)) \end{bmatrix}.$$

Evaluating Equations 8, 9 at all times $t = t_1, \ldots, t_N$ yields

$$\begin{bmatrix} | & | & & | \\ \dot{\mathbf{X}}_1 & \dot{\mathbf{X}}_2 & \cdots & \dot{\mathbf{X}}_n \\ | & | & & | \end{bmatrix} = \dot{\mathbf{X}} \approx \mathbf{\Theta}(\mathbf{X})\mathbf{\Xi}$$

$$= \begin{bmatrix} | & & | \\ \theta_1(\mathbf{X}) & \cdots & \theta_p(\mathbf{X}) \\ | & & | \end{bmatrix} \begin{bmatrix} | & | & \cdots & | \\ \xi_1 & \xi_2 & \cdots & \xi_n \\ | & | & \cdots & | \end{bmatrix} \qquad (11)$$

Sparse regression techniques, such as LASSO (Tibshirani, 1996) or sequential thresholded least-squares (Brunton et al., 2016), can then solve the overdetermined system (Equation 11) for the sparse $\mathbf{\Xi}$. This provides an approximation to the governing equation as in Equation 10.

While SINDy can be directly applied to Examples 1–3 to discover closed moment systems based on simulated PDE data (Equation 1), our specific interest lies in the following:

1. Given time series data of moments with a known closed moment system (e.g., Examples (1) and (2)), can we robustly discover the governing dynamics using a correct but potentially oversized dictionary $\mathbf{\Theta}(\mathbf{x})$, such as polynomials up to a high degree?
2. If a system is not closed for a chosen set of moments, but closure exists after a proper coordinate transformation

(e.g., selecting $\mathbf{x} = [I_2, V_1, K, J]^{\top}$ in Ex. 3), what insights can we gain by directly applying SINDy to time series data from this "incorrect" set of moments?
3. For more general systems where analytical closure does not exist, can data-driven methods provide a "good enough" numerical approximation to predict the future evolution of the moment system, effectively serving as a principled reduced-order description of the underlying PDE?

These questions will be addressed in Section 4. Of particular interest is the second point, where we demonstrate that, in certain cases, we can gain insight into the appropriate transformation to close the system, even if the moment system is not closed under the originally selected variables. In the following section, we discuss a more principled strategy to discover such transformations in a data-driven fashion.

3.2 Data-driven discovery of coordinate transformations for moment system closure

To illustrate the idea, we focus on Example 3, where the initially selected moments are $\mathbf{x} = [I_2, V_1, K, J]^{\top}$, and an analytical closure exists *only after* a coordinate transformation, $\mathbf{y} = \mathbf{A}^{\top}\mathbf{x}$,

$$\mathbf{y} = \begin{bmatrix} I_2 \\ V_1 \\ K+J \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} I_2 \\ V_1 \\ K \\ J \end{bmatrix} = \mathbf{A}^{\top}\mathbf{x},$$

Remark 1. Note that the transformation matrix $\mathbf{A} \in \mathbb{R}^{4 \times 3}$ is not unique. For any full-rank matrix $\mathbf{P} \in \mathbb{R}^{3 \times 3}$, the moment system remains closed under the transformation $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}^{\top}\mathbf{x}$, with $\tilde{\mathbf{A}} = \mathbf{AP}$.

Assume we aim to discover $\mathbf{A}$ purely from simulated PDE data. We propose the following strategy: Let $\mathbf{X} \in \mathbb{R}^{N \times 4}$ be the state matrix for the original coordinate $\mathbf{x} = [I_2, V_1, K, J]^{\top}$, sampled at $t_1, \ldots, t_N$, as defined by Equation 8. Let $\mathbf{y} = \mathbf{A}^{\top}\mathbf{x}$ be the new coordinate, and the associated new state matrix $\mathbf{Y} \in \mathbb{R}^{N \times 3}$ becomes

$$\mathbf{Y} = \begin{bmatrix} -\, \mathbf{y}(t_1)^{\top}\, - \\ -\, \mathbf{y}(t_2)^{\top}\, - \\ \vdots \\ -\, \mathbf{y}(t_N)^{\top}\, - \end{bmatrix} = \begin{bmatrix} -\, \mathbf{x}(t_1)^{\top}\, - \\ -\, \mathbf{x}(t_2)^{\top}\, - \\ \vdots \\ -\, \mathbf{x}(t_N)^{\top}\, - \end{bmatrix} \mathbf{A} = \mathbf{XA}$$

We can then solve for $\mathbf{A}$ through the following optimization.

$$\min_{\mathbf{A} \in \mathbb{R}^{4 \times 3}} \; \min_{\mathbf{\Xi} \in \mathbb{R}^{p \times 3}} \left\| \frac{d}{dt}(\mathbf{XA}) - \mathbf{\Theta}(\mathbf{XA}) \cdot \mathbf{\Xi} \right\|_F^2 + \mu \|\mathbf{\Xi}\|_1, \qquad (12)$$

$$\text{s.t. } \mathbf{A}^{\top}\mathbf{A} = \mathbf{I}_{3 \times 3}$$

where $\frac{d}{dt}(\mathbf{XA})$ is the derivative matrix, Equation 8, associated with the new state matrix $\mathbf{Y} = \mathbf{XA}$, $\| \cdot \|_F$ and $\| \cdot \|_1$ are the Frobenius norm and l^1 norm, respectively, defined for any $m \times n$ matrix $\mathbf{B} = (b_{ij})$ as follows:

$$\|\mathbf{B}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |b_{ij}|^2},$$

$$\|\mathbf{B}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |b_{ij}|.$$

In addition, $\mu \geq 0$ is a non-negative weight, and $\Theta(\mathbf{XA}) = \Theta(\mathbf{Y}) \in \mathbb{R}^{N \times p}$ is the library matrix, Equation 8, of a chosen library $\Theta(\mathbf{y})$ on the new coordinate $\mathbf{y} = \mathbf{A}^\top \mathbf{x}$.

The idea behind Equation 12 is very simple: we search for the transformation matrix $\mathbf{A} \in \mathbb{R}^{4 \times 3}$ such that the dynamics under the new coordinate $\mathbf{y} = \mathbf{A}^\top \mathbf{x}$ can be parsimoniously represented from the library $\Theta(\mathbf{y})$. The weight $\mu \geq 0$ controls the sparsity-promoting L_1-regularization, and the constraint $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_{3 \times 3}$ prevents the trivial solution $\mathbf{A} = \mathbf{0}$. The set of $\mathbf{A}$ satisfying the constraint $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_{3 \times 3}$ is called the *Stiefel manifold* (Edelman et al., 1998; Absil et al., 2008).

Remark 2. When the weight $\mu = 0$, even with the Stiefel manifold constraint $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_{3 \times 3}$, the solution to Equation 12 is not unique. Just like Remark 1, if $(\mathbf{A}^\star, \Xi^\star)$ is a solution, then for any orthogonal matrix $\mathbf{O} \in \mathrm{O}(3)$, the pair $(\tilde{\mathbf{A}}^\star, \tilde{\Xi}^\star)$ is also a solution, where

$$\tilde{\mathbf{A}}^\star = \mathbf{A}^\star \mathbf{O}, \quad \tilde{\Xi}^\star = \mathbf{O}^\top \Xi^\star \mathbf{O}$$

To solve Equation 12, we use alternating optimization, iteratively optimizing $\mathbf{A}$ and Ξ while keeping the other fixed. Specifically, when $\mathbf{A}$ is fixed, solving for Ξ reduces to a LASSO problem. Conversely, when Ξ is fixed, the problem becomes a Stiefel manifold optimization with a smooth objective function, which can be efficiently solved using methods such as those presented in (Oviedo and Dalmau, 2019; Xiao et al., 2020; Liu et al., 2021). To ensure the algorithm remains unbiased, we implement an annealing strategy for the hyperparameter μ. Initially, μ is kept constant for `Iter_scheduled` iterations. Following this period, μ is reduced by half every α iterations. Refer to Algorithm 1 for the pseudocode to solve problem Equation 12.

Required: maxIter: Total number of iterations.
    `Iter_scheduled`: Number of initial iterations during which μ remains constant.
    α: Number of iterations after which μ is halved.
    Random initializations: $\mathbf{A}^{(0)} \in \mathbb{R}^{4 \times 3}$ on the Stiefel manifold and $\Xi^{(0)} \in \mathbb{R}^{p \times 3}$.
  Output: $\mathbf{A}_{\text{out}}, \Xi_{\text{out}}$
1 **for** k in $[1, 2, \cdots, \text{maxIter}]$ **do**
2   **if** $k > \textit{Iter\_scheduled}$ and k is a multiple of α **then**
3      $\mu \leftarrow \mu/2$;
4   **end**
5    $\Xi^{(k)} \leftarrow \arg\min_\Xi \left\| \frac{d}{dt}\left(\mathbf{XA}^{(k-1)}\right) - \Theta\left(\mathbf{XA}^{(k-1)}\right) \cdot \Xi \right\|_F^2 + \mu\|\Xi\|_1$, solved by LASSO.
6    $\mathbf{A}^{(k)} \leftarrow \arg\min_{\mathbf{A}^\top \mathbf{A} = \mathbf{I}} \left\| \frac{d}{dt}(\mathbf{XA}) - \Theta(\mathbf{XA}) \cdot \Xi^{(k-1)} \right\|_F^2$, solved by Stiefel optimization.
7 **end**
8 $\mathbf{A}_{\text{out}} \leftarrow \mathbf{A}^{(\text{maxIter})}$
9 $\Xi_{\text{out}} \leftarrow \Xi^{(\text{maxIter})}$

Algorithm 1. Data-driven discovery of coordinate transformation.

4 Numerical results

In this section, we present numerical results on data-driven closure of moment systems (Examples 1–3) using the methodologies described in Section 3. To obtain the data, we first numerically solve the PDE Equation 1 with periodic boundary conditions and various ICs using an extended 4th-order Runge-Kutta method, the exponential integrator (ETDRK4) (see (Kassam and Trefethen, 2005) for a detailed explanation). The time series of the moments is subsequently extracted by numerical spatial integration according to Equations 2–5, with spatial derivatives computed using pseudo-spectral Fourier methods. We note once again that should the numerical data be replaced by "experimental" ones from a given physical process, the procedure can still be applied. Unless otherwise noted, the time series are evaluated

at $N = 16,000$ uniformly spaced times $t = t_1, \ldots, t_N$, where $t_i = i\Delta t$ and $\Delta t = 0.0025$.

4.1 Examples with analytical moment closure

We first examine Examples (1) and (2), where analytical (linear) closure exists for the chosen moments $\mathbf{x} = [I_1, V_0]$ (Example 1) and $\mathbf{x} = [I_2, V_1, K]$ (Example 2).

4.2 Example 1

Let the selected moments be $\mathbf{x} = [I_1, V_0]$. We construct the data matrices $\mathbf{X}^{(0)} = [\mathbf{I}_1^{(0)}, \mathbf{V}_0^{(0)}] \in \mathbb{R}^{N \times 2}$ and $\mathbf{X}^{(1)} = [\mathbf{I}_1^{(1)}, \mathbf{V}_0^{(1)}] \in \mathbb{R}^{N \times 2}$ from numerically solving the PDE Equation 1 with the ICs:

$$u^{(1)}(x, 0) = \pi^{-1/4} \exp\left(-\frac{1}{2}(x - 5)^2\right),$$

$$u^{(2)}(x, 0) = \frac{1}{2}\mathrm{sech}^2(x - 5).$$

We apply SINDy to this system using monomial libraries of degree up to $n \in \mathbb{N}$, $\Theta_{\deg \leq n}(\mathbf{x})$, as defined in Equation 9. We find that as long as $n \leq 2$, SINDy applied to either data matrix $\mathbf{X}^{(0)}$ or $\mathbf{X}^{(1)}$ discovers the dynamics nearly perfectly. However, as we gradually expand the library $\Theta_{\deg \leq n}(\mathbf{x})$ by increasing n, SINDy applied to either $\mathbf{X}^{(0)}$ or $\mathbf{X}^{(1)}$ *individually* typically produces erroneous dynamics, regardless of how carefully the sparsity-promoting parameter is tuned. A typical negative result from SINDy applied to $\mathbf{X}^{(0)}$ with $n = 3$ is presented in the Supplementary Material. This outcome is expected, as increasing the library size makes the problem more ill-posed, leading SINDy to overfit the data and produce incorrect dynamics.

To address the overfitting issue, we can enlarge the dataset by concatenating the data matrices $\mathbf{X}^{(0)}$ and $\mathbf{X}^{(1)}$ vertically, forming $\mathbf{X} = [\mathbf{X}^{(0)\top}, \mathbf{X}^{(1)\top}]^\top \in \mathbb{R}^{2N \times 2}$. When applying SINDy to this new data matrix $\mathbf{X}$ (considering boundary issues when taking finite differences), SINDy can now discover the correct dynamics that match Equation 6, even with a much larger library $\Theta_{\deg \leq n}(\mathbf{x})$ for n up to 16. For example, when $n = 16$, the output ODE from SINDy reads

$$\begin{cases} \dfrac{dI_1}{dt} = 1.000V_0, \\[2mm] \dfrac{dV_0}{dt} = -1.000I_1, \end{cases}$$

where the coefficients are rounded to three decimal places. This suggests the potential usefulness of concatenating different time series, especially in cases where one may not be familiar with the order of the relevant closure.

4.3 Example 2

For Example 2 with the selected moments $\mathbf{x} = [I_2, V_1, K]$, our findings are similar to those in Section 4.2. When applying SINDy to a data matrix from a single IC, SINDy discovers erroneous dynamics for the quadratic dictionary $\Theta_{\deg \leq 2}$. However, using larger data

matrices from multiple ICs, SINDy can once again accurately identify the correct dynamics even with the larger dictionaries. Representative negative and positive results are presented in the Supplementary Material, similar to the previous example.

4.4 Examples where closure exists after coordinate transformations

We now turn to Example 3, where the selected moments are $\mathbf{x} = [I_2, V_1, K, J]$, and the moment system only closes after a coordinate transformation. Specifically, we consider the nonlinearity $g(\rho, t) = \rho^2$ in Equation 1, which satisfies the condition in Example 3. We collect the moment time series data by numerically solving the PDE with the following four distinct ICs, Equations 13–16:

$$u^{(1)}(x, 0) = \pi^{-1/4} \exp\left(-\frac{1}{2}(x-5)^2\right), \tag{13}$$

$$u^{(2)}(x, 0) = 1.88 \exp\left(-\frac{1}{2}(x-5)^2\right), \tag{14}$$

$$u^{(3)}(x, 0) = 1.88\left(\exp\left(-\frac{1}{2}(x-5)^2\right) + \exp\left(-(x-2)^2\right)\right), \tag{15}$$

$$u^{(4)}(x, 0) = 1.88(\cos(2x) + \sin(2x))\exp\left(-x^2\right). \tag{16}$$

We consider the following questions:

- What insights can we gain by directly applying SINDy to this system with the selected moments $\mathbf{x} = [I_2, V_1, K, J]$, where a closure does not exist?
- Can the method described in Section 3.2 correctly identify the coordinate transformation that closes the moment system?

We note that *normalized* moment time series are used as input for SINDy, as a way of incorporating feature scaling. This is an important aspect that ensures that all the relevant quantities are considered on "equal footing," when the sparse regression step takes place.

4.5 SINDy with linear library $\Theta_{\deg=1}(\mathbf{x})$

We begin by applying SINDy with a linear library $\Theta_{\deg=1}(\mathbf{x})$ to the moment time series data from IC Equation 14. The resulting equations are:

$$\begin{cases} \dfrac{dI_2}{dt} = 1.000V_1, \\[2mm] \dfrac{dV_1}{dt} = -2.000I_2 + 4.000K + 3.998J, \\[2mm] \dfrac{dK}{dt} = -0.569V_1, \\[2mm] \dfrac{dJ}{dt} = 0.069V_1. \end{cases} \tag{17}$$

The coefficients are rounded to three decimal places. The first four panels of Figure 1 compare the "ground-truth" time-evolution data of $[I_2, V_1, K, J]$ obtained from PDE integration (red curves) and the definition of these quantities through Equations 2–5 with those from integrating the SINDy-predicted ODEs Equation 17 (blue curves). While the SINDy-predicted time evolution of $[I_2, V_1]$ closely matches

the ground truth, there is a *significant discrepancy* for the moments $[K, J]$ in Figure 1. Accordingly, the SINDy-predicted dynamics is not accurate for the original coordinates $\mathbf{x} = [I_2, V_1, K, J]$.

Nevertheless, interestingly, if we add the ODE for K with that for J in Equation 17, which corresponds to the *correct* coordinate transformation for closure, we obtain:

$$\begin{cases} \dfrac{dI_2}{dt} = 1.000V_1, \\[2mm] \dfrac{dV_1}{dt} = -2.000I_2 + 4.000K + 3.998J, \\[2mm] \dfrac{d(K+J)}{dt} = -0.500V_1. \end{cases} \tag{18}$$

Equation 18 closely matches the ground-truth moment system Equation 7. The last panel of Figure 1 shows that the predicted time evolution of $K + J$ aligns perfectly with the ground truth, even though SINDy does not accurately recover those of K and J *individually*. This finding is intriguing because, even when SINDy is applied to the original coordinates $[I_2, V_1, K, J]$ without closure, it "strategically chooses" to sacrifice accuracy for K and J but indirectly captures the correct dynamics when considering the proper coordinate transformation $E = K + J$. Moreover, a similar pattern is consistently observed when applying SINDy to the moments simulated from the other three ICs. The results are shown in the Supplementary Material. We have also provided an analogous example for 2D NLS equations. For details, please refer to Section 3 of the Supplementary Material. Indeed, this clearly illustrates that the methodology presented herein is not in any fundamental way limited to 1D NLS models, but can be extended to higher dimensional ones, as may be deemed desirable.
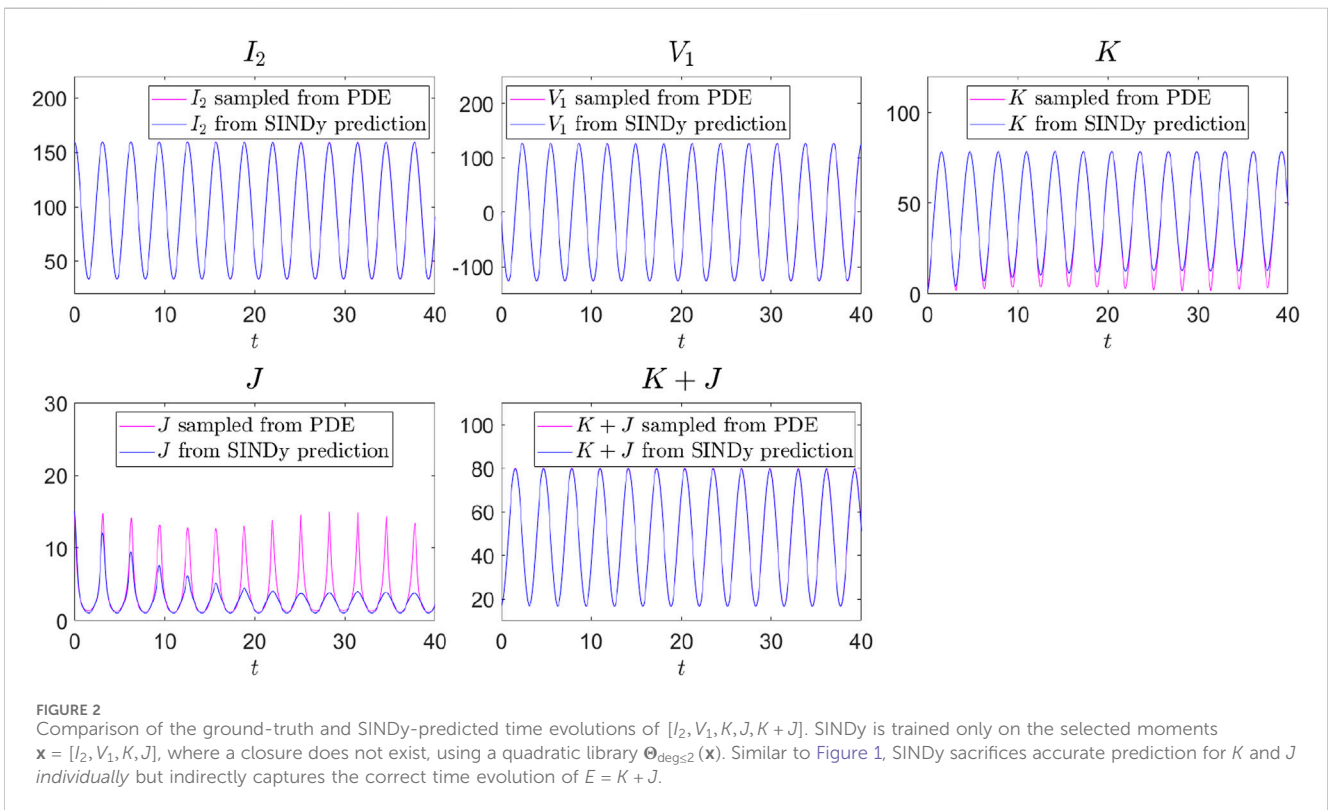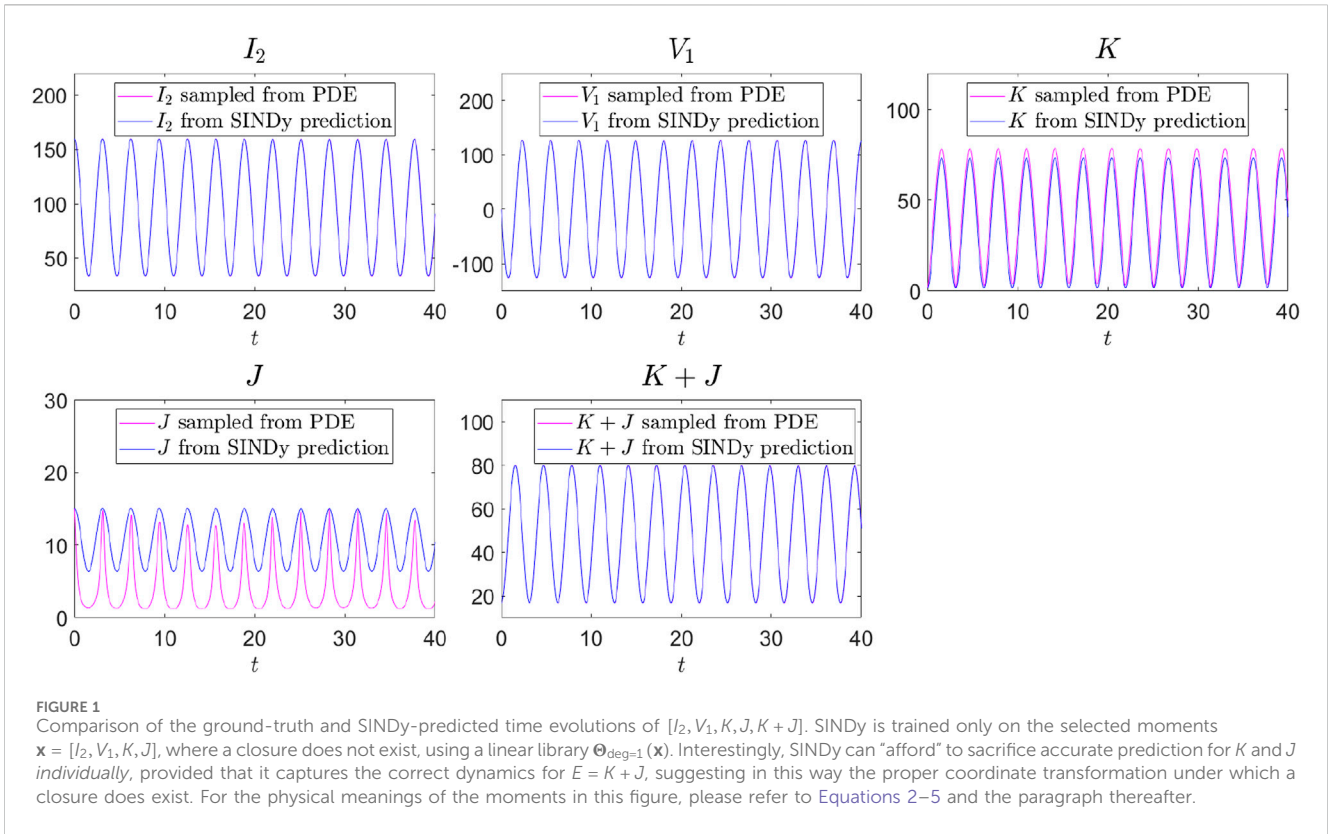
4.6 SINDy with quadratic library $\Theta_{\deg\leq2}(\mathbf{x})$

Next, we investigate the performance of SINDy with an expanded quadratic library $\Theta_{\deg\leq2}(\mathbf{x})$ applied to the moment time series with the same IC Equation 14. The predicted ODEs now become:

$$\begin{cases} \dfrac{dI_2}{dt} = 1.000V_1, \\[2mm] \dfrac{dV_1}{dt} = -2.000I_2 + 4.000K + 3.998J, \\[2mm] \dfrac{dK}{dt} = -0.174V_1 - 0.003V_1K - 0.083V_1J, \\[2mm] \dfrac{dJ}{dt} = -0.002I_2V_1 + 0.082V_1J. \end{cases} \tag{19}$$

As before, if we add the predicted dynamics of K and J in Equation 19, we obtain:

$$\begin{cases} \dfrac{dI_2}{dt} = 1.000V_1, \\[2mm] \dfrac{dV_1}{dt} = -2.000I_2 + 4.000K + 3.998J, \\[2mm] \dfrac{d(K+J)}{dt} = -0.174V_1 - 0.003V_1K - 0.002I_2V_1 - 0.001V_1J. \end{cases} \tag{20}$$

FIGURE 1
Comparison of the ground-truth and SINDy-predicted time evolutions of $[I_2, V_1, K, J, K + J]$. SINDy is trained only on the selected moments $\mathbf{x} = [I_2, V_1, K, J]$, where a closure does not exist, using a linear library $\Theta_{\deg=1}(\mathbf{x})$. Interestingly, SINDy can "afford" to sacrifice accurate prediction for K and J *individually*, provided that it captures the correct dynamics for $E = K + J$, suggesting in this way the proper coordinate transformation under which a closure does exist. For the physical meanings of the moments in this figure, please refer to Equations 2–5 and the paragraph thereafter.

FIGURE 2
Comparison of the ground-truth and SINDy-predicted time evolutions of $[I_2, V_1, K, J, K + J]$. SINDy is trained only on the selected moments $\mathbf{x} = [I_2, V_1, K, J]$, where a closure does not exist, using a quadratic library $\Theta_{\deg\leq2}(\mathbf{x})$. Similar to Figure 1, SINDy sacrifices accurate prediction for K and J *individually* but indirectly captures the correct time evolution of $E = K + J$.

Unlike the previous case with $\Theta_{\deg=1}(\mathbf{x})$, the SINDy-predicted governing Equation 20 using the expanded library $\Theta_{\deg\leq2}(\mathbf{x})$ after the (theoretically motivated) coordinate transformation $E = K + J$ still fails to match the ground truth ODE Equation 7 and remains unclosed (i.e., the third equation in Equation 20 cannot be written using only I_2, V_1, and $E = K + J$). Specifically, the coefficient of V_1

in Equation 20 is −0.174, deviating substantially from that of −0.5 in the ground truth. Although the coefficients for the additional terms $(V_1K, I_2V_1,$ and $V_1J)$ are relatively small, their presence hinders the accurate recovery of the correct coefficient for V_1.

Figure 2 compares the ground-truth time evolutions of $[I_2, V_1, K, J, E = K + J]$ obtained from PDE integration with their SINDy-predicted dynamics after integrating Equations 19, 20. Similar to Figure 1, SINDy with the quadratic library $\Theta_{\deg\leq2}(\mathbf{x})$ applied to the unclosed moments $[I_2, V_1, K, J]$ sacrifices the exact recovery of the time evolution of J and K, but indirectly captures the correct time evolution of $E = J + K$. However, unlike the linear library case in Section 4.5, SINDy with the quadratic library can yield the correct *time evolution* of $E = K + J$ but *fails* to uncover the correct *governing equation* for E [Equation 20] due to the expanded library size, leading to overfitting issues discussed above.

Here, we see a key complication within SINDy that has been also present in works such as (Champion et al., 2019) and especially (Bakarji et al., 2022), namely the methodology is likely to result in ODE models that are proximal to the theoretically expected ones, but not identical to them. This, in turn, may result in a nontrivial error outside of the training set. Especially when the library at hand is "richer" than the terms expected to be present, unfortunately, it does not generically seem that the reduced, theoretically expected model will be discovered. Rather, our results suggest that it is possible that the additional "wealth" of the libraries used can be leveraged to approximate the data via different (nonlinear) dependent variable combinations.

As an even more problematic example, applying SINDy with a quadratic library to the moment time series generated from another IC Equation 16 results in a predicted ODE system that not only fails to match the ground truth, even after the theoretically suggested coordinate transformation $K + J$, but also causes the ODE system to blow up in finite time. For a detailed discussion, we refer the interested reader to the Supplementary Material.

4.7 Stiefel optimization for discovering coordinate transformations

Next, we test the methodology from Section 3.2 to discover the coordinate transformation needed to close the moment system.

Case 1: $\mu = 0$. We first consider the case where the hyperparameter μ in Equation 12 is set to $\mu = 0$, i.e., we do not promote sparsity in the matrix Ξ. We use the linear library $\Theta(\mathbf{x}) = \Theta_{\deg=1}(\mathbf{x})$, and set the maximum number of iterations to maxIter = 150 in Algorithm 1. The algorithm produces the following outputs

$$\mathbf{A}_{out} \approx \begin{bmatrix} -0.973 & 0.179 & 0.144 \\ -0.169 & -0.982 & 0.083 \\ -0.111 & -0.040 & -0.697 \\ -0.111 & -0.040 & -0.697 \end{bmatrix}, \quad \Xi_{out} \approx \begin{bmatrix} -0.024 & -1.076 & 0.005 \\ 1.016 & 0.470 & -0.540 \\ 0.913 & 5.778 & -0.446 \end{bmatrix}.$$
(21)

The coefficients are rounded to three decimal places. On the other hand, the "ground-truth" coordinate transformation $\mathbf{A}_{gt}$ (after normalization to satisfy the Stiefel manifold constraint) and the corresponding dynamics $\mathbf{A}_{gt}$ according to Equation 7 are:

$$\mathbf{A}_{gt} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \dfrac{1}{\sqrt{2}} \\ 0 & 0 & \dfrac{1}{\sqrt{2}} \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.707 \\ 0 & 0 & 0.707 \end{bmatrix},$$

$$\Xi_{gt} = \begin{bmatrix} 0 & -2 & 0 \\ 1 & 0 & -\dfrac{1}{2\sqrt{2}} \\ 0 & 4\sqrt{2} & 0 \end{bmatrix} \approx \begin{bmatrix} 0 & -2 & 0 \\ 1 & 0 & -0.354 \\ 0 & 5.657 & 0 \end{bmatrix} \quad (22)$$

At first glance, the predicted solution Equation 21 appears different from the ground truth in Equation 22. However, after a further change of coordinates using

$$\mathbf{O} = \begin{bmatrix} -0.973 & -0.169 & -0.156 \\ 0.179 & -0.982 & -0.0565 \\ 0.144 & 0.0830 & -0.986 \end{bmatrix} \in O(3),$$

we have

$$\mathbf{A}_{out}\mathbf{O} = \begin{bmatrix} 1.000 & -0.000 & 0.000 \\ 0.000 & 1.000 & -0.000 \\ 0.000 & -0.000 & 0.707 \\ 0.000 & 0.000 & 0.707 \end{bmatrix} \approx \mathbf{A}_{gt},$$

$$\mathbf{O}^\top\Xi_{out}\mathbf{O} = \begin{bmatrix} 0.000 & -2.000 & 0.000 \\ 1.000 & 0.000 & -0.354 \\ 0.000 & 5.657 & 0.000 \end{bmatrix} \approx \Xi_{gt}.$$

Hence, according to Remark 2, $(\mathbf{A}_{out}, \Xi_{out})$ and $(\mathbf{A}_{gt}, \Xi_{gt})$ are equivalent solutions, and our method successfully predicts the correct coordinate transformation $\mathbf{A}_{out}$ to close the moment system. However, as expected, due to $\mu = 0$, the linear combination matrix Ξ_{out} of the transformed dictionary is not sparse.

Case 2: $\mu > 0$. We next explore the case when $\mu \neq 0$ in Equation 12. We use the same linear library $\Theta(\mathbf{x}) = \Theta_{\deg=1}(\mathbf{x})$, and μ is initially set to $\mu = 1$. In Algorithm 1, we set maxIter = 1000, Iter_scheduled = 400, and $\alpha = 20$. The algorithm returns

$$\tilde{\mathbf{A}}_{out} \approx \begin{bmatrix} -1.000 & 0.000 & 0.030 \\ 0.000 & -0.100 & 0.002 \\ -0.021 & -0.001 & -0.707 \\ -0.021 & -0.001 & -0.707 \end{bmatrix}, \quad \tilde{\Xi}_{out} \approx \begin{bmatrix} 0.000 & -1.830 & 0.003 \\ 1.000 & 0.011 & -0.383 \\ 0.000 & 5.714 & -0.010 \end{bmatrix}.$$
(23)

Similarly, after a further change of coordinates using:

$$\tilde{\mathbf{O}} = \begin{bmatrix} -1.000 & 0.000 & -0.030 \\ 0.000 & -1.000 & -0.002 \\ 0.030 & 0.002 & -1 \end{bmatrix} \in O(3),$$

we again have

$$\tilde{\mathbf{A}}_{out}\tilde{\mathbf{O}} = \begin{bmatrix} 1.000 & -0.000 & 0.000 \\ 0.000 & 1.000 & -0.000 \\ 0.000 & -0.000 & 0.707 \\ 0.000 & 0.000 & 0.707 \end{bmatrix} \approx \mathbf{A}_{gt},$$

$$\tilde{\mathbf{O}}^\top\tilde{\Xi}_{out}\tilde{\mathbf{O}} = \begin{bmatrix} 0.000 & -2.000 & 0.000 \\ 1.000 & 0.000 & -0.354 \\ 0.000 & 5.657 & 0.000 \end{bmatrix} \approx \Xi_{gt}.$$
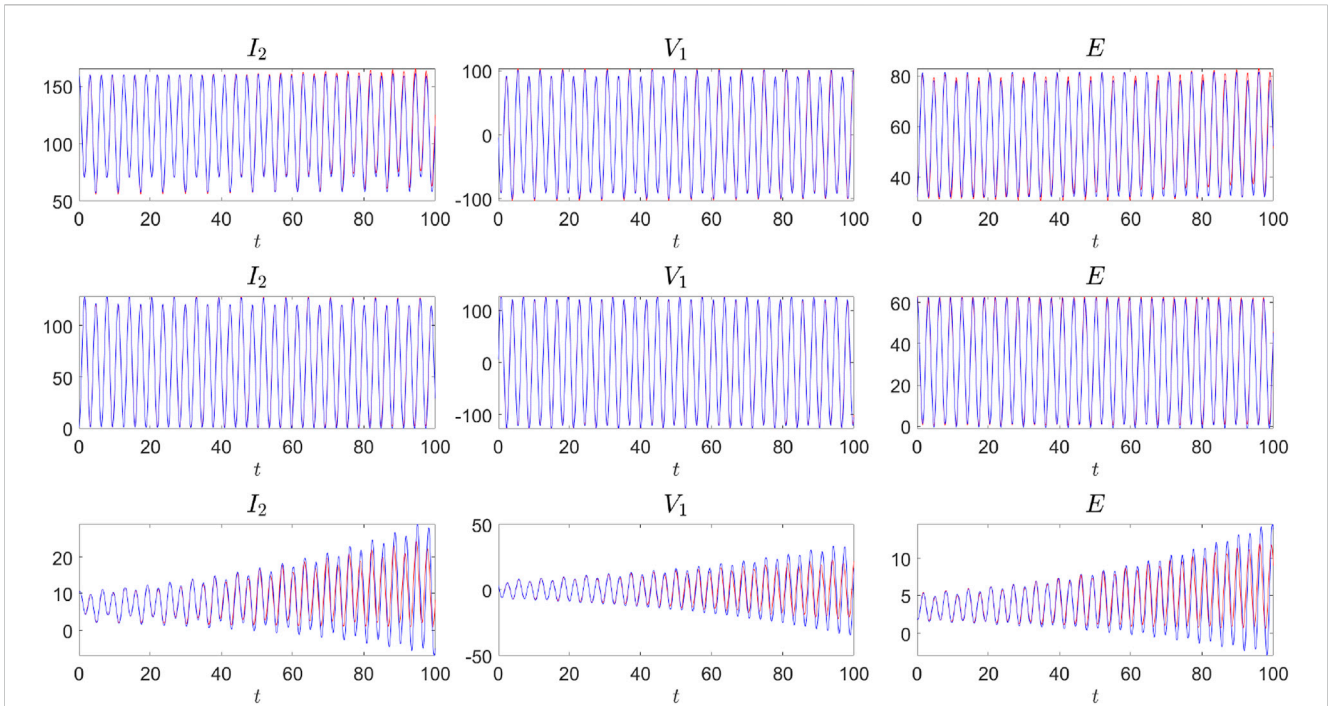
FIGURE 3
Comparison between the ground-truth time evolution (in red) and the SINDy-predicted evolution (in blue). Training data were obtained by integrating the NLS equation up to $T = 20$ using the initial conditions of Equations 25–27. In all cases (displayed in three different rows), the SINDy-predicted moment dynamics closely match the ground truth, well past the time frame of the training data, offering a satisfactory reduced-order description of the underlying PDE through the corresponding moment systems.

Note that the result $(\tilde{\mathbf{A}}_{out}, \tilde{\Xi}_{out})$ in Equation 23 for $\mu > 0$ is much closer to the ground truth $(\mathbf{A}_{gt}, \Xi_{gt})$ in Equation 22, compared to $(\mathbf{A}_{out}, \Xi_{out})$ in Equation 21 with $\mu = 0$. This demonstrates that a positive μ with annealed optimization not only discovers the correct coordinate transformation to close the system but also achieves a sparser solution for Ξ, reducing the number of terms on the right-hand side of the moment ODE system.

4.8 An example of unclosed moment system

Finally, we present a case without analytical closure. Our "data-driven closure" aims to provide an accurate, reduced-order description of the PDE by approximating the evolution of the moment systems. In principle, this is the type of problem that we are aiming for, namely the discovery of potential moment closures when these may not be analytically available; the examples presented previously are valuable benchmarks to raise the complications that may emerge when one seeks to use this type of methodology in systems where the answer may be unknown and what credibility one may wish to assign to the obtained results.

Specifically, we consider an NLS Equation 1 with a *time-dependent* nonlinearity:

$$g(\rho, t) = (\sin(t) + 2)|\rho|^2. \tag{24}$$

Notice that such time-dependent nonlinearities are well-known for some time in atomic physics settings (Donley et al., 2001; Staliunas et al., 2002) (and continue to yield novel insights to

this day (Shagalov and Friedland, 2024)) and similar dynamical scenarios have been considered in nonlinear optics (Centurion et al., 2006; Zhang et al., 2021).

Moment systems with such nonlinearity will not close to the best of our knowledge, so we aim to numerically approximate form of the dynamics of the moments $\mathbf{x} = [I_2, V_1, E]$, where $E = K + J$. We consider the following three ICs:

$$u_1(x, 0) = 1.88 \exp\left(-\frac{1}{2}(x - 5)^2\right), \tag{25}$$

$$u_2(x, 0) = 1.88(\cos(2x) + \sin(2x))\exp(-x^2), \tag{26}$$

$$u_3(x, 0) = \exp(-0.1x^2)\exp(0.1ix^2). \tag{27}$$

Notably, the IC Equation 27 includes a quadratic phase, motivated by the quadratic phase approximation (QPA) ansatz for NLS equations discussed by (Pérez-García et al., 2007). It also includes regular, smooth localized initial conditions, as well as one involving Fourier mode oscillations, modulated by the Gaussian term. We expect the SINDy-predicted dynamics to vary with the different ICs of Equations 25–27, hence the relevant choices.

Due to the periodic nature of the nonlinearity g in Equation 24, we expect the moment system to be non-autonomous and exhibit an oscillatory pattern. To capture this, we introduce the following two "auxiliary moments":

$$\begin{cases} H = \sin(t) + 2, \\ Q = \cos(t) + 2. \end{cases}$$

Naturally, one can observe that these are inspired by the nature of $g(\rho, t)$. However, one can envision the use of Fourier modes even

if the mathematical model was not known or if the data stemmed from experimental observations.

We then apply SINDy with a linear library $\Theta_{\deg=1}(\tilde{\mathbf{x}})$ to the expanded moment systems $\tilde{\mathbf{x}} = [\mathbf{x}, H, Q] = [I_2, V_1, E, H, Q]$. The time series data of the moments are collected by integrating the PDE up to $T = 20$. The exact SINDy-predicted ODEs for all three ICs are presented in the Supplementary Material.

To test the accuracy of the predicted moment systems, we integrate the SINDy-predicted ODEs into the future, up to $T = 100$. Figure 3 compares the ground-truth and SINDy-predicted time evolutions for the three ICs Equations 25–27. In all cases, the model closely matches the ground truth up to $T = 100$. Indeed, this is well past the training time of $T = 20$ and thus provides a satisfactory reduced-order description of the underlying PDE effective moment dynamics. Thus, despite the potential shortcomings of the method which we tried to present in an unbiased fashion in case examples where the analytical theory helps assess them, we still find it to be a worthwhile tool to consider. I.e., from a data-driven perspective, it can be seen to potentially provide an effective, low-dimensional dynamical representation of the associated high-dimensional PDE dynamics.

5 Conclusion and future directions

This study explores a data-driven approach to identifying moment equations in nonlinear Schrödinger models. This paves the way more generally towards the use of similar methods in nonlinear PDEs which may feature similar wave phenomena. We applied the relevant sparse regression/optimization methodology aiming to rediscover known analytical closures (progressively extending considerations to more complex settings), addressed overfitting by augmenting datasets with multiple initial conditions, and identified suitable coordinate transformations for systems requiring them so as to bring forth the reduced or analytically tractable form of the dynamics. Additionally, we demonstrated that our approach could provide a reduced-order description of systems without analytical closures by approximating the evolution of the moment systems. Our findings show that this data-driven method can capture complex dynamics in NLS models and offer insights for various physical applications, possibly well past the training time used for the data-driven methods.

Future work will focus on extending our method to more complex PDEs and exploring its applicability to other types of nonlinear dynamical high-dimensional models, such as, e.g., the ones we mentioned in the context Fisher-KPP models and their applications to brain tumor dynamics, as analyzed, e.g., in (Belmonte-Beitia et al., 2014). Another possible avenue is to, instead of recovering the moment systems through *numerical differentiation* of the moment time series (as done in SINDy), leverage *numerical integration* into future time of a suitably augmented system. This approach, similar to Neural ODEs (Chen et al., 2018) and shooting methods, can help avoid producing predicted ODEs that blow up in finite time which SINDy may produce (see details in Section 4.2.2 and the Supplementary Material). Additionally, we plan to develop techniques to identify *nonlinear* coordinate transformations that

can close the moment system, further enhancing the applicability of our method. Lastly, one can envision such classes of techniques for obtaining additional reduced features of solitary waves, such as data-driven variants of the variational approximation (Malomed, 2002), or data-driven models of soliton interaction dynamics (Manton, 1979; Kevrekidis et al., 2004; Ma et al., 2016).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

SY: Formal Analysis, Investigation, Methodology, Software, Writing–original draft. SC: Formal Analysis, Investigation, Methodology, Software, Writing–original draft. WZ: Conceptualization, Funding acquisition, Methodology, Writing–review and editing. PK: Conceptualization, Funding acquisition, Methodology, Supervision, Writing–review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphot.2024.1444993/full#supplementary-material

References

Ablowitz, M. (2011). *Nonlinear dispersive waves, asymptotic analysis and solitons*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511998324

Ablowitz, M., and Clarkson, P. (1991). "Solitons, nonlinear evolution equations, and inverse scattering," in *149 of London math. Soc. Lecture note series*. Cambridge: Cambridge U. Press.

Ablowitz, M., Prinari, B., and Trubatch, A. (2004). *Discrete and continuous nonlinear schrödinger systems*. Cambridge: Cambridge University Press.

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Bakarji, J., Champion, K. P., Kutz, J. N., and Brunton, S. L. (2022). Discovering governing equations from partial measurements with deep delay autoencoders. *CoRR abs/2201*, 05136. doi:10.1098/rspa.2023.0422

Belmonte-Beitia, J., Calvo, G. F., and Pérez-García, V. M. (2014). Effective particle methods for Fisher–Kolmogorov equations: theory and applications to brain tumor dynamics. *Commun. Nonlinear Sci. Numer. Simul.* 19, 3267–3283. doi:10.1016/j.cnsns.2014.02.004

Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016). *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, 3932–3937.

Centurion, M., Porter, M. A., Kevrekidis, P. G., and Psaltis, D. (2006). Nonlinearity management in optics: experiment, theory, and simulation. *Phys. Rev. Lett.* 97, 033903. doi:10.1103/PhysRevLett.97.033903

Champion, K., Lusch, B., Kutz, J. N., and Brunton, S. L. (2019). Data-driven discovery of coordinates and governing equations. *Proc. Natl. Acad. Sci.* 116, 22445–22451. doi:10.1073/pnas.1906995116

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). Neural ordinary differential equations. *Adv. neural Inf. Process. Syst.* 31. doi:10.48550/arXiv.1806.07366

De Koster, P., and Wahls, S. (2024). Data-driven identification of the spectral operator in AKNS Lax pairs using conserved quantities. *Wave Motion* 127, 103273. doi:10.1016/j.wavemoti.2024.103273

Donley, E. A., Claussen, N. R., Cornish, S. L., Roberts, J. L., Cornell, E. A., and Wieman, C. E. (2001). Dynamics of collapsing and exploding Bose–Einstein condensates. *Nature* 412, 295–299. doi:10.1038/35085500

Edelman, A., Arias, T. A., and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Analysis Appl.* 20, 303–353. doi:10.1137/s0895479895290954

Feliu-Faba, J., Fan, Y., and Ying, L. (2020). Meta-learning pseudo-differential operators with deep neural networks. *J. Comput. Phys.* 408, 109309. doi:10.1016/j.jcp.2020.109309

García-Ripoll, J., and Pérez-García, V. (1999). The moment method in general nonlinear Schrödinger equations

Hasegawa, A., and Kodama, Y. (1995). *Solitons in optical communications*. Oxford: Clarendon Press.

Infeld, E., and Rowlands, G. (2000). *Nonlinear waves, solitons and chaos*. Cambridge: Cambridge University Press.

Karjanto, N. (2024). Modeling wave packet dynamics and exploring applications: a comprehensive guide to the nonlinear Schrödinger equation. *Mathematics* 12, 744. doi:10.3390/math12050744

Karniadakis, G., Kevrekidis, I., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nat. Rev. Phys.* 3, 422–440. doi:10.1038/s42254-021-00314-5

Kassam, A.-K., and Trefethen, L. N. (2005). Fourth-order time-stepping for stiff PDEs. *SIAM J. Sci. Comput.* 26, 1214–1233. doi:10.1137/s1064827502410633

Kevrekidis, P. G., Frantzeskakis, D. J., and Carretero-González, R. (2015). *The defocusing nonlinear schrödinger equation*. Philadelphia: SIAM.

Kevrekidis, P. G., Khare, A., and Saxena, A. (2004). Solitary wave interactions in dispersive equations using Manton's approach. *Phys. Rev. E* 70, 057603. doi:10.1103/PhysRevE.70.057603

Kivshar, Y. S., and Agrawal, G. P. (2003). *Optical solitons: from fibers to photonic crystals*. Academic Press. doi:10.1016/B978-0-12-410590-4.X5000-1

Kono, M., and Skorić, M. (2010). *Nonlinear physics of plasmas*. Heidelberg: Springer-Verlag.

Krippendorf, S., Lüst, D., and Syvaeri, M. (2021). Integrability ex machina. *Fortschritte Phys.* 69, 2100057. doi:10.1002/prop.202100057

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., et al. (2021). "Fourier neural operator for parametric partial differential equations," in International Conference on Learning Representations.

Liu, X., Xiao, N., and Xiang Yuan, Y. (2021). A penalty-free infeasible approach for a class of nonsmooth optimization problems over the Stiefel manifold. *J. Sci. Comput.* 99, 30. doi:10.1007/s10915-024-02495-4

Liu, Z., Madhavan, V., and Tegmark, M. (2022). Machine learning conservation laws from differential equations. *Phys. Rev. E* 106, 045307. doi:10.1103/physreve.106.045307

Liu, Z., and Tegmark, M. (2021). Machine learning conservation laws from trajectories. *Phys. Rev. Lett.* 126, 180604. doi:10.1103/PhysRevLett.126.180604

Liu, Z., and Tegmark, M. (2022). Machine learning hidden symmetries. *Phys. Rev. Lett.* 128, 180201. doi:10.1103/PhysRevLett.128.180201

Lu, L., Meng, X., Mao, Z., and Karniadakis, G. (2021). DeepXDE: a deep learning library for solving differential equations. *SIAM Rev.* 63, 208–228. doi:10.1137/19m1274067

Ma, M., Navarro, R., and Carretero-González, R. (2016). Solitons riding on solitons and the quantum Newton's cradle. *Phys. Rev. E* 93, 022202. doi:10.1103/PhysRevE.93.022202

Malomed, B. A. (2002). Progress in optics (Elsevier). *Var. methods nonlinear fiber Opt. Relat. fields* 43, 69–191. doi:10.1016/S0079-6638(02)80026-9

Manton, N. (1979). An effective Lagrangian for solitons. *Nucl. Phys. B* 150, 397–412. doi:10.1016/0550-3213(79)90309-2

Oviedo, H., and Dalmau, O. (2019). "A scaled gradient projection method for minimization over the Stiefel manifold," in *Advances in soft computing*. Editors L. Martínez-Villaseñor, I. Batyrshin, and A. Marín-Hernández (Cham: Springer International Publishing), 239–250.

Pérez-García, V., Torres, P., and Montesinos, G. (2007). The method of moments for nonlinear Schrödinger equations: theory and applications. *SIAM J. Appl. Math.* 67, 990–1015. doi:10.1137/050643131

Pethick, C. J., and Smith, H. (2002). *Bose–einstein condensation in Dilute Gases*. Cambridge, United Kingdom: Cambridge University Press.

Pitaevskii, L., and Stringari, S. (2003). *Bose-Einstein condensation*. Oxford: Oxford University Press.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707. doi:10.1016/j.jcp.2018.10.045

Schaeffer, H. (2017). Learning partial differential equations via data discovery and sparse optimization. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 473, 20160446. doi:10.1098/rspa.2016.0446

Shagalov, A. G., and Friedland, L. (2024). Autoresonant generation of solitons in Bose-Einstein condensates by modulation of the interaction strength. *Phys. Rev. E* 109, 014201. doi:10.1103/PhysRevE.109.014201

Staliunas, K., Longhi, S., and de Valcárcel, G. J. (2002). Faraday patterns in Bose-Einstein condensates. *Phys. Rev. Lett.* 89, 210406. doi:10.1103/PhysRevLett.89.210406

Sulem, C., and Sulem, P. (1999). *The nonlinear Schrödinger equation: self-focusing and wave collapse*. New York: Springer.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Xiao, N., Liu, X., and Yuan, Y.-x. (2020). A class of smooth exact penalty function methods for optimization problems with orthogonality constraints. *Optim. Methods Softw.* 37, 1205–1241. doi:10.1080/10556788.2020.1852236

Zhang, S., Fu, Z., Zhu, B., Fan, G., Chen, Y., Wang, S., et al. (2021). Solitary beam propagation in periodic layered Kerr media enables high-efficiency pulse compression and mode self-cleaning. *Light Sci. and Appl.* 10, 53. doi:10.1038/s41377-021-00495-9

Zhu, W., Zhang, H.-K., and Kevrekidis, P. G. (2023). Machine learning of independent conservation laws through neural deflation. *Phys. Rev. E* 108, L022301. doi:10.1103/PhysRevE.108.L022301