Chapter 2

Weak form-based data-driven modeling

Computationally efficient and noise robust equation learning and parameter inference

David M. Bortz*, **Daniel A. Messenger, and April Tran**

Department of Applied Mathematics, University of Colorado, Boulder, CO, United States

**Corresponding author: e-mail address: david.bortz@colorado.edu*

Contents

Abstract

Recent developments in the field of data-driven modeling have led to the advancement of weak form-based methodologies in scientific machine learning. This class of methods offers several compelling advantages, including high computational efficiency and high noise robustness. In this chapter, we present an overview of the weak form approach as well as discuss several categories of applications including equation discovery, constitutive parameter inference, and reduced order modeling. In particular, we illustrate the performance on several benchmark examples for ordinary, partial, and stochastic differential equations.

Handbook of Numerical Analysis, Volume 25, ISSN 1570-8659, https://doi.org/10.1016/bs.hna.2024.05.002

Keywords
Equation learning, Data-driven modeling, Weak form, WSINDy, WENDy, WLaSDI

MSC Codes
37M10, 62-07, 62FXX, 62JXX, 65L09, 65M32, 65R99, 92-08

1 Introduction

The *weak form* of a differential equation is created by multiplying both sides by a smooth function ϕ and integrating over a domain of interest. Historically, this idea originated from the fact that frequently physical conservation laws can intuitively be cast as integral equations. Moreover, simple integral (i.e., $\phi \equiv 1$) and variational formulations consistently allow for easier analysis and simulation of a broad class of models (including those with shocks and other solution discontinuities). In the twentieth century, Sobolev was the first to suggest that ϕ be (what Friedrichs later named a *mollifier*) a compactly supported C^∞ function which integrates to one, while Schwartz rigorously recasts the classical notion of a function acting on a point to one acting on a measurement structure or *test function* (ϕ) (Schwartz, 1950). Modern computational approaches based on the weak form (such as the finite element method (FEM) for solving an equation) originated with the work of many researchers including Argyris, Courant, Friedrichs, Galerkin, Hrennikoff, Oganesya, and others (see Liu et al., 2022, for an overview of the history). Lax and Milgram then built on these efforts to establish a theoretical foundation by proving the existence of weak solutions (in a Hilbert Space) to certain classes of PDEs (Lax and Milgram, 1955).

While it is clear that the weak form is used widely in the analysis and simulation of differential equations, recent advances suggest that there is an incredible potential for novel methods applied to statistical and machine learning problems. In particular, weak form versions of sparse regression-based equation learning, direct parameters estimation, and reduced order modeling can offer advantages in terms of noise robustness and computational speedups over both conventional methods and black-box neural network approaches.

The overall goal of this handbook chapter is to both introduce the basics of weak form system identification as well as communicate the breadth of possible applications. Accordingly, in §2, we provide an overview of the Weak form Sparse Identification of Nonlinear Dynamics (WSINDy) method (§2.2) in the context of equation discovery for ODEs (§2.3) and PDEs (§2.4), summarizing the contributions from Messenger and Bortz (2021a,b). Motivated by the fact that weak form-based learning is demonstrably more robust to noise than strong form-based approaches, in §3 we present a novel theoretical explanation for the improved performance (under certain assumptions).

In §4 we explore how a direct parameter estimation method substantially benefits from casting the equation in the weak form. We build on our Weak form Estimation of Nonlinear Dynamics (WENDy) applied to ODEs (Bortz

et al., 2023) and we extend the method to PDEs and SDEs. Lastly, in §5 we demonstrate how WENDy can complement existing SVD- and autoencoder-based reduced order modeling (ROM) techniques by substantially enhancing the discovery of the latent space dynamics from noisy data.

2 Weak form-based equation discovery

In this section, we discuss a class of methods that use sparse regression to learn governing equations directly from data. The idea is to consider a large library of potential functional forms for terms on the right side of an evolution equation. As described below, the library and the measured data are combined to create a sparse regression problem, the answer to which reveals the subset of terms on the right side which best relate the data to its time derivative.

We will begin by describing the first widely successful equation discovery method and then discuss how to cast this problem as one in which the equation is learned in the weak form. We will then present examples using conventional ODEs and PDEs. However, we also note that the examples presented here are only a subset of models and applications to which we have applied the weak form. We have considered integro-differential equation models (Messenger et al., 2022b), online learning (Messenger et al., 2022a), and coarse-graining of stochastic differential equations (Messenger and Bortz, 2022b; Messenger et al., 2023).

2.1 The sparse identification of nonlinear dynamics (SINDy) method for learning governing equations

To begin, let us consider an evolution equation system

$$\frac{d}{dt}\mathbf{u}(t) = \mathbf{F}(\mathbf{u}(t)), \quad \mathbf{u}(0) = \mathbf{u}_0 \in \mathbb{R}^d, \quad 0 \le t \le T, \tag{1}$$

with data $\mathbf{U} \in \mathbb{R}^{M\times d}$ observed at M timepoints $\mathbf{t} = (t_1, \ldots, t_M)^T$ by

$$\mathbf{U}_{md} = \mathbf{u}_i(t_m) + \epsilon_{mi}, \qquad m \in [M],\ i \in [d].$$

Note that we use the bracket notation $[M] := \{1, \ldots, M\}$ and that the variable $\epsilon \in \mathbb{R}^{M\times d}$ is a matrix of i.i.d. measurement noise. The goal of the equation learning problem is to discover the dynamics (1) from the measurements $\mathbf{U}$.

In 2016, Brunton, Proctor, and Kutz published an article in the Proceedings of the National Academy (Brunton et al., 2016) introducing the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm. This algorithm has been successful used in learning parsimonius nonlinear dynamics when (among other considerations) noise is small. This framework assumes that the ith element of the function $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ in (1) is

$$\mathbf{F}_i(\mathbf{u}(t)) = \sum_{j=1}^{J} \mathbf{w}^\star_{ji}\, f_j(\mathbf{u}(t)) \tag{2}$$

for some known family of functions $(f_j)_{j\in[J]}$ and a coefficient weight matrix $\mathbf{w}^\star \in \mathbb{R}^{J\times d}$ which is mostly zeros, i.e., sparse.

To learn the model, first a data matrix is built $\Theta(\mathbf{U}) \in \mathbb{R}^{M\times J}$ by substituting the data directly into the equation

$$[\Theta(\mathbf{U})]_{(m,j)} = f_j(\mathbf{U}_m), \qquad \mathbf{U}_m := (\mathbf{U}_{m1}, \ldots, \mathbf{U}_{md}),$$

so that the candidate functions are directly evaluated at the measured data. Model discovery is thus recast as solving for a sparse $\mathbf{w}^\star$ by minimizing (subject to a sparsity penalty) the L^2 norm of the residual

$$\left\| \dot{\mathbf{U}} - \Theta(\mathbf{U})\,\widehat{\mathbf{w}} \right\|_2, \tag{3}$$

where $\dot{\mathbf{U}}$ is the numerical time derivative of the data $\mathbf{U}$. In SINDy, sequential-thresholding least squares on the parameters $\mathbf{w}$ are then used to arrive at a sparse solution.

2.2 Weak form SINDy (WSINDy)

While SINDy has been used successfully in a wide range of areas, it is well known that using a simple finite difference approximation for the derivatives will lead to poor performance in the presence of even modest amounts of noise. There have been several efforts to address the problem using methods such as total-variation regularized derivatives (Brunton et al., 2016), linear multistep derivative approximations (Chen, 2023), and ensembling (Fasel et al., 2022) (among others), each with varying degrees of success.

Weak-form extensions to SINDy have been independently discovered by several groups (Gurevich et al., 2019; Messenger and Bortz, 2021a,b; Pantazis and Tsamardinos, 2019; Wang et al., 2019) over the past few years. Although our group was not the first to propose a weak-form methodology, we have investigated its use for equation learning in a wide range of model structures and applications including: ODEs (Messenger and Bortz, 2021b), PDEs (Messenger and Bortz, 2021a), interacting particle systems of the first (Messenger and Bortz, 2022b) and second (Messenger et al., 2022b) order, and online streaming (Messenger and Bortz, 2022b). As mentioned below, we have also studied the theoretical convergence properties for WSINDy in the continuum data limit (Messenger and Bortz, 2022a). This work led to specification of a broad class of models for which the asymptotic limit of continuum data can overcome any noise level to produce both an accurately learned equation and a correct parameter estimate (see Messenger and Bortz, 2022a, for more information).

To describe WSINDy, we begin by recalling that for any smooth test function $\phi : \mathbb{R} \to \mathbb{R}$ and interval $(a, b) \subset [0, T]$, Eq. (1) admits the weak formulation

$$\begin{aligned} &\phi(b)\mathbf{u}(b) - \phi(a)\mathbf{u}(a) - \int_a^b \dot{\phi}(t)\,\mathbf{u}(t)\,dt = \int_a^b \phi(t)\,\mathbf{F}(\mathbf{u}(t))\,dt, \\ &0 \le a < b \le T. \end{aligned} \tag{4}$$

With $\phi = 1$, we arrive at the integral equation of the dynamics explored in Schaeffer and McCalla (2017). If we instead take ϕ to be nonconstant and compactly supported in (a, b), we arrive at

$$-\int_a^b \dot{\phi}(t)\,\mathbf{u}(t)\,dt = \int_a^b \phi(t)\,\mathbf{F}(\mathbf{u}(t))\,dt. \tag{5}$$

Assuming a representation is of the form (2), we then define the generalized residual $\mathcal{R}(\mathbf{w}; \phi)$ for a given test function ϕ and a set of candidate terms $(f_j)_{j\in[J]}$ and with data $\mathbf{U}$ as follows:

$$\mathcal{R}(\mathbf{w}; \phi) := \int_a^b \left(\dot{\phi}(t)\,\mathbf{U}(t) + \phi(t) \left(\sum_{j=1}^{J} \mathbf{w}_j\, f_j(\mathbf{U}(t)) \right) \right) dt. \tag{6}$$

Clearly, with $\mathbf{w} = \mathbf{w}^\star$ and $\mathbf{U} = \mathbf{u}(t)$ we have $\mathcal{R}(\mathbf{w}; \phi) = 0$ for all ϕ compactly-supported in (a, b); however, $\mathbf{U}$ is a finite dimensional vector of data, hence the integral in (6) must be approximated numerically. Measurement noise then presents a further barrier to accurate identification of $\mathbf{w}^\star$.

In Algorithm 1, we now state the Weak SINDy method in full generality. We propose a generalized least squares approach with approximate covariance matrix Σ. Below we derive a particular choice of Σ which utilizes the action of the test functions $(\phi_k)_{k\in[K]}$ on the data $\mathbf{y}$. Sequential-thresholding on the weight coefficients $\mathbf{w}$ with thresholding parameter λ is used to enforce sparsity, where $\lambda \leq \min_{\mathbf{w}^\star \neq 0} |\mathbf{w}^\star|$ is necessary for recovery. Lastly, an ℓ_2-regularization term with coefficient γ is included for problems involving rank deficiency.

WSINDy has proven to be successful in discovering governing equations for a wide class of differential equations. In this section, we will illustrate the application to ODEs, PDEs, and the mean field limit equation of a specific class of SDEs. We note that the ODE example in Section 2.3 is a straightforward application of the formulae in Algorithm 1. However, the later examples for PDEs (Section 2.4) need more exposition and notation to be fully described.

2.3 WSINDy for ordinary differential equations

Recall the canonical Lorenz system with a set of parameters known to induce solutions with chaotic dynamics

$$\begin{aligned} \dot{x} &= 10(y - x) \\ \dot{y} &= x(28 - z) - y \\ \dot{z} &= xy - \tfrac{8}{3}z. \end{aligned}$$

We solve this ODE using a variable order, variable stepsize solver based on numerical differentiation formulas of orders 1 to 5, as implemented in Matlab®'s ode15s with absolute and relative tolerances of 10^{-12}.

Algorithm 1: Weak form Sparse Identification of Nonlinear Dynamics (WSINDy)

input : Data $\{\mathbf{t}, \mathbf{U}\}$, Candidate functions $(f_j)_{j\in[J]}$, Test Functions $(\phi_k)_{k\in[K]}$, Variance Σ, Regularization (λ, γ)

output: Parameter Estimate $\widehat{\mathbf{w}}$

```
// Construct matrix of trial gridfunctions
```
$\Theta(\mathbf{U}) \leftarrow \left[f_1(\mathbf{U}) \,|\, \ldots \,|\, f_J(\mathbf{U})\right]$

```
// Construct integration matrices
```
$[\boldsymbol{\Phi}]_{(k,m)} \leftarrow \Delta t \phi_k(t_m)$

$[\dot{\boldsymbol{\Phi}}]_{(k,m)} \leftarrow \Delta t \dot{\phi}_k(t_m)$

```
// Compute Gram matrix and right side vector
```
$[\mathbf{G}]_{(k,j)} = \langle \phi_k, f_j(\mathbf{U}) \rangle$

$[\mathbf{b}]_{(k,i)} = -\langle \phi_k', \mathbf{U}_i \rangle$

```
// Solve the following generalized least squares problem
   with ℓ2-regularization using
// sequential thresholding with parameter λ to enforce
   sparsity.
```
$\widehat{\mathbf{w}} = \text{argmin}_{\mathbf{w}} \left\{(\mathbf{G}\mathbf{w} - \mathbf{b})^T \Sigma^{-1} (\mathbf{G}\mathbf{w} - \mathbf{b}) + \gamma^2 \|\mathbf{w}\|_2^2\right\}$

To apply WSINDy to the ODE equation recovery problem, a test function must be specified. In our initial efforts to learn ODEs (Messenger and Bortz, 2021b), we empirically discovered that a smooth piecewise polynomial provides excellent results

$$\phi(t, r, p) := \left(1 - \left(\frac{t}{r}\right)^2\right)^p.$$

In Messenger and Bortz (2021b), we proposed a strategy for choosing the compact support radius r and the power p. While we have discovered more accurate and sophisticated strategies for tuning these parameters (see Section 2.4), there are several general principles. First, in general, the power p will need to be much larger than the value needed to perform integration by parts. Second, the radius depends upon the dynamics and typically larger radii work better. Notably, in the absence of noise and using a large radius and a highly smooth test function, WSINDy recovers the parameters to machine precision-level accuracy.

Depicted in Fig. 1 are the result of using WSINDy to discover the governing Lorenz equations from data with 10% additive Gaussian noise. In the upper left subfigure, the (high numerical accuracy) solution is in black and the sampled data are in red. In the upper right subfigure, the black is the same as in the upper

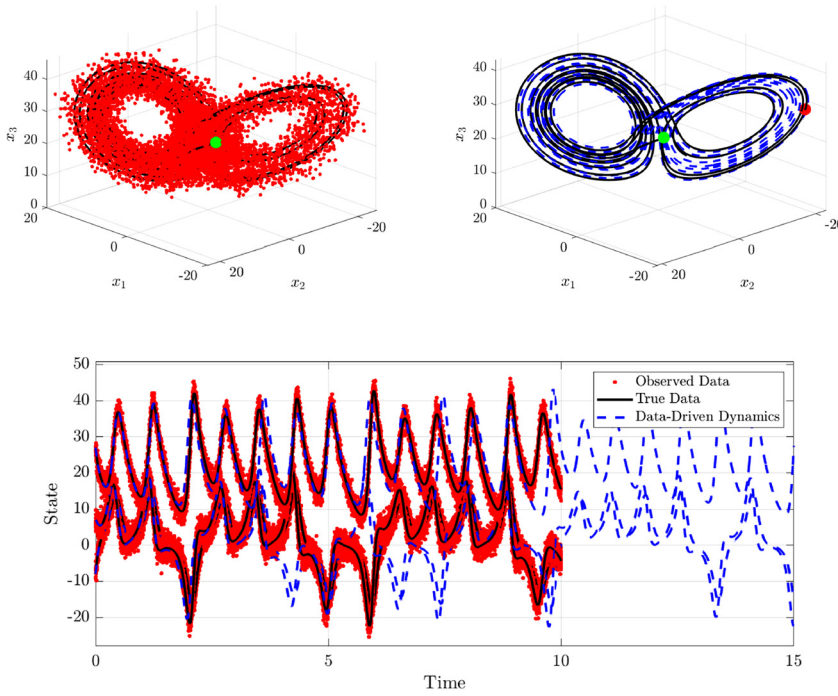


FIGURE 1 Lorenz system with $(x_0, y_0, z_0) = (-8,\ 7,\ 27)^T$. All correct terms were identified with an error in the weights of $E_2(\widehat{w}t) = 0.0084$ and trajectory error $\mathcal{E}(\widehat{w}) = 0.56$. The large trajectory error is expected due to the chaotic nature of the solution. Using data up until $t = 1.5$ (first 1500 timepoints) the trajectory error is 0.027. Figure from Messenger and Bortz (2021b).

left and the blue dotted curve is a simulation using the discovered parameters. As is clear in the bottom subfigure, the simulated solution is close to the highly accurate one, but given that the system is chaotic, it is notable that the discovered solution provides a nearly perfect match until about 4 time units.

For more ODE examples including Lotka-Volterra, Van der Pol, Duffing, etc., we direct the interested reader to Messenger and Bortz (2021b).

2.4 WSINDy for partial differential equations

We are now ready to discuss the formulation of WSINDy for PDEs. We assume that the set of multiindices $(\alpha^s)_{s\in[S]}$ together with α^0 enumerates the set of possible true differential operators that govern the evolution of u and that $(g_s)_{s\in[S]} \subset \text{span}(f_j)_{j\in[J]}$ where the family of functions $(f_j)_{j\in[J]}$ (referred to as the *trial functions*) is known beforehand. This enables us to write down a general class of PDEs as

$$\mathscr{D}^{\alpha^0} u = \sum_{s=1}^{S} \sum_{j=1}^{J} \mathbf{w}^\star_{(s-1)J+j} \mathscr{D}^{\alpha^s} f_j(u), \tag{7}$$

so that discovery of the correct PDE is reduced to a finite-dimensional problem of recovering the true vector of coefficients $\mathbf{w}^\star \in \mathbb{R}^{SJ}$, which is assumed to be sparse. We emphasize that a wide variety of PDEs can be written in the form (7) including inviscid Burgers, Korteweg-de Vries, Kuramoto-Sivashinsky, nonlinear Schrödinger's, Sine-Gordon, a class of reaction-diffusion systems, and Navier-Stokes.

To convert the PDE into its weak form, we multiply Eq. (7) by a smooth *test function* $\psi(x,t)$, compactly-supported in $\Omega \times (0,T)$, and integrate over the spacetime domain,

$$\left\langle \psi,\ \mathscr{D}^{\alpha^0} u \right\rangle = \sum_{s=1}^{S} \sum_{j=1}^{J} \mathbf{w}^\star_{(s-1)J+j} \left\langle \psi,\ \mathscr{D}^{\alpha^s} f_j(u) \right\rangle,$$

where the L^2-inner product is defined $\langle \psi, f \rangle := \int_0^T \int_\Omega \psi(x,t) f(x,t)\, dx dt$. Using the compact support of ψ and Fubini's theorem, we then integrate by parts as many times as necessary to arrive at the following weak form of the dynamics:

$$\left\langle (-1)^{|\alpha^0|} \mathscr{D}^{\alpha^0} \psi,\ u \right\rangle = \sum_{s=1}^{S} \sum_{j=1}^{J} \mathbf{w}^\star_{(s-1)J+j} \left\langle (-1)^{|\alpha^s|} \mathscr{D}^{\alpha^s} \psi,\ f_j(u) \right\rangle, \tag{8}$$

where $|\alpha^s| := \sum_{d=1}^{D+1} \alpha_d^s$ is the order of the multiindex.[1] Using an ensemble of test functions $(\psi_k)_{k\in[K]}$, we then discretize the integrals in (8) with $f_j(u)$ replaced by $f_j(\mathbf{U})$ (i.e. evaluated at the observed data $\mathbf{U}$) to arrive at the linear least squares problem

$$\min_w \|\mathbf{b} - \mathbf{G}\mathbf{w}\|_2^2$$

defined by

$$\begin{cases} \mathbf{b}_k = \left\langle (-1)^{|\alpha^0|} \mathscr{D}^{\alpha^0} \psi_k,\ \mathbf{U} \right\rangle, \\ \mathbf{G}_{k,(s-1)J+j} = \left\langle (-1)^{|\alpha^s|} \mathscr{D}^{\alpha^s} \psi_k,\ f_j(\mathbf{U}) \right\rangle, \end{cases} \tag{9}$$

where $\mathbf{b} \in \mathbb{R}^K$, $\mathbf{G} \in \mathbb{R}^{K \times SJ}$ and $\mathbf{w} \in \mathbb{R}^{SJ}$ are using the inner product both in the sense of a continuous and exact integral in (8) and a numerical approximation in (9) which depends on a chosen quadrature rule.[2]

[1] For example, with $\mathscr{D}^{\alpha^s} = \frac{\partial^{2+1}}{\partial x^2 \partial y}$, integration by parts occurs twice with respect to the x-coordinate and once with respect to y, so that $|\alpha^s| = 3$ and $(-1)^{|\alpha^s|} = -1$.

[2] In all cases in this chapter, we use the trapezoidal rule, see Messenger and Bortz (2021b) for a discussion.

For the WSINDy algorithm, the key pieces of the algorithm are (i) the choice of reference test function ψ, (ii) the method of a sparsification, (iii) the method of regularization, (iv) selection of convolution query points $\{(\mathbf{x}_k, t_k)\}_{k\in K}$, and (v) the model library. Full guidance for choices for each of these hyperparameters (in the case of PDEs) is provided in Messenger and Bortz (2021a).

Here we will simply note that the choice of test function is central to the performance and it is still an open question as to the optimal functions for different scenarios. For the illustrating examples below, we use a piecewise polynomial

$$\phi(v) = \begin{cases} C(v-a)^p(b-v)^q & a < v < b, \\ 0 & \text{otherwise}, \end{cases} \tag{10}$$

where $p, q \geq 1$ and v is a (time or space) variable. The normalization

$$C = \frac{1}{p^p q^q}\left(\frac{p+q}{b-a}\right)^{p+q}$$

ensures that $\|\phi\|_\infty = 1$. For ease of computation, multiplicative test functions are used, e.g.,

$$\psi(t, \mathbf{x}) = \phi_0(t)\prod_{i=1}^{D}\phi_i(x_i)$$

for a D-dimensional space. Examples of these test functions in one time and one space dimension are depicted in Fig. 2. Lastly, we note that the test functions defined here are functions of both time and space (instead of just of time as in Section 2.3).

Table 1 lists the PDEs used to demonstrate the performance of WSINDy. We numerically solved[3] these equations and then created artificial data by adding i.i.d. Gaussian noise with variance σ^2 to each data point. The value of σ is constructed to depend on the root mean squares (RMS) of the data, i.e., $\sigma := \sigma_{NR}\|\mathbf{U}^\star\|_{RMS}$ and we refer to σ_{NR} as the *noise ratio*.

There are standard metrics for measuring the accuracy of parameter estimates, such as the relative ℓ^2 error

$$E_2(\widehat{w}) := \frac{\|\widehat{w} - \mathbf{w}^\star\|_{RMS}}{\|\mathbf{w}^\star\|_{RMS}}, \tag{11}$$

which we will employ. For the equation discovery problem, we will use the *true positivity ratio* introduced in Lagergren et al. (2020) and defined by

$$\text{TPR}(\widehat{w}) = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}, \tag{12}$$

[3] Details on the numerical methods and boundary conditions used to simulate each PDE can be found in Messenger and Bortz (2021a).

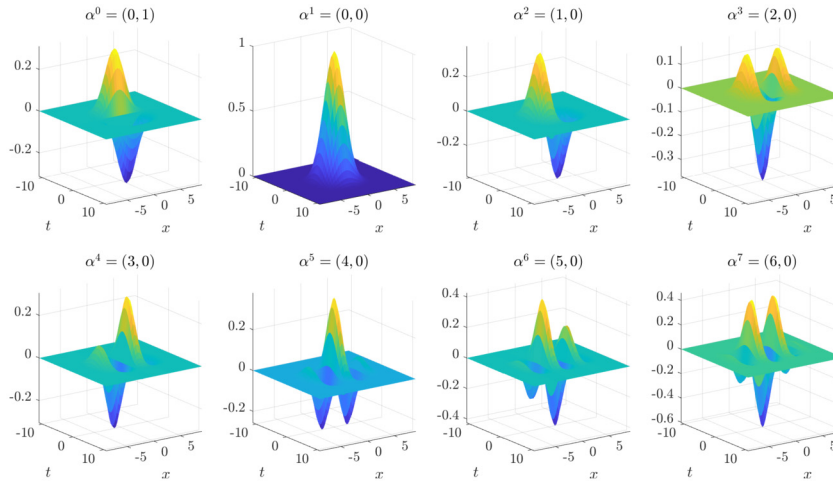


FIGURE 2 Plots of reference test function ψ and partial derivatives $\mathscr{D}^{\alpha^s}\psi$ used for identification of the Kuramoto-Sivashinsky equation. The upper left plot shows $\partial_t\psi$, the bottom right shows $\partial_x^6\psi$. See Tables 1–2 for more details. Reproduced from Messenger and Bortz (2021a).

TABLE 1 PDEs used in numerical experiments, written in the form identified by WSINDy. Domain specification and boundary conditions are given in Messenger and Bortz (2021a).

Inviscid Burgers (IB)	$\partial_t u = -\frac{1}{2}\partial_x(u^2)$
Korteweg-de Vries (KdV)	$\partial_t u = -\frac{1}{2}\partial_x(u^2) - \partial_{xxx}u$
Kuramoto-Sivashinsky (KS)	$\partial_t u = -\frac{1}{2}\partial_x(u^2) - \partial_{xx}u - \partial_{xxxx}u$
Nonlinear Schrödinger (NLS)	$\begin{cases}\partial_t u = \frac{1}{2}\partial_{xx}v + u^2v + v^3\\ \partial_t v = -\frac{1}{2}\partial_{xx}u - uv^2 - u^3\end{cases}$
Anisotropic Porous Medium (PM)	$\partial_t u = (0.3)\partial_{xx}(u^2) - (0.8)\partial_{xy}(u^2) + \partial_{yy}(u^2)$
Sine-Gordon (SG)	$\partial_{tt}u = \partial_{xx}u + \partial_{yy}u - \sin(u)$
Reaction-Diffusion (RD)	$\begin{cases}\partial_t u = \frac{1}{10}\partial_{xx}u + \frac{1}{10}\partial_{yy}u - uv^2 - u^3 + v^3 + u^2v + u\\ \partial_t v = \frac{1}{10}\partial_{xx}v + \frac{1}{10}\partial_{yy}v + v - uv^2 - u^3 - v^3 - u^2v\end{cases}$
2D Navier-Stokes (NS)	$\partial_t\omega = -\partial_x(\omega u) - \partial_y(\omega v) + \frac{1}{100}\partial_{xx}\omega + \frac{1}{100}\partial_{yy}\omega$

where TP is the number of correctly identified nonzero coefficients, FN is the number of coefficients falsely identified as zero, and FP is the number of coefficients falsely identified as nonzero. Identification of the true model results in a TPR of 1, while identification of half of the correct nonzero terms and no falsely identified nonzero terms results in TPR of 0.5.

TABLE 2 Computational efficiency of WSINDy for learning the listed PDEs. The $\widetilde{\mathbf{G}}$ column reports the size of the matrix used with in the sparse regression where the tilde denotes that the data has been scaled to improve the computation stability. The next column indicates the condition number after the rescaling. The last column shows the start-to-finish walltime with all computations in serial measured on a laptop with an 8-core Intel i7-2670QM CPU. Notably, none of the computations take more than 75 seconds. Moreover, these walltimes are independent of the noise since the same algorithm is being used regardless of the σ_{NR} level.

PDE	$\widetilde{\mathbf{G}}$	$\kappa(\widetilde{\mathbf{G}})$	Walltime (sec)
IB	784×43	1.4×10^6	0.12
KdV	1443×43	3.2×10^6	0.39
KS	1806×43	3.7×10^3	0.24
NLS	1804×190	1.2×10^5	2.5
NS	3872×50	8.2×10^2	12
PM	4608×65	2.4×10^4	16
SG	$13{,}000 \times 73$	1.3×10^4	29
RD	$11{,}638 \times 181$	4.5×10^3	75

For each system in Table 1 and a range of noise levels $\sigma_{NR} \in [0, 1]$ we ran[4] WSINDy with 200 realizations of noise and average the results.

Fig. 3 depicts the TPR for all the PDEs investigated. Notably, for inviscid Burgers, Korteweg-de Vries, Kuramoto-Sivashinsky and Sine-Gordon, the average TPR stays above 0.95 even for $\sigma_{NR} = 1$. The average TPR for the nonlinear Schrödinger and porous medium equations stays above 0.95 until 50% noise, after which identification of the correct monomial nonlinearity is not as reliable. For NLS, this is a drastic improvement over previous studies (Rudy et al., 2017), especially considering the large library of 190 terms used.

Fig. 4 depicts the E_2 error in the recovered coefficients for $\sigma_{NR} \in [0, 1]$. Intriguingly, for a noise level of 10%, the relative error is typically 10% or less in almost all cases. This is surprising as in many cases, conventional methods (depending on a comparison of data with numerical solutions) yield relative errors on the order of or higher than the σ_{NR} from the data (see §4.1.1 and Bortz et al., 2023).

[4] Computations were run on a 16-core Intel Xeon 5218 CPU node in the Colorado Research Computing condo system.

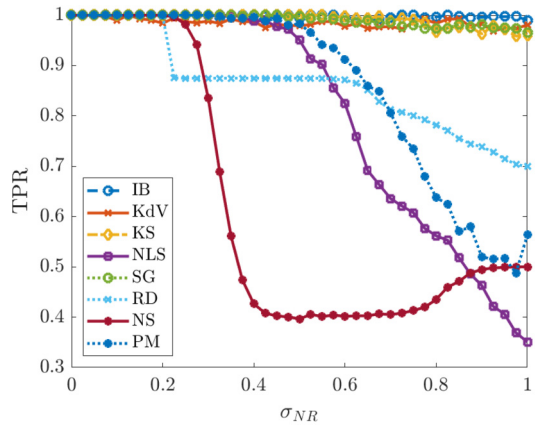


FIGURE 3 Average TPR (true positivity ratio, defined in (12)) for each of the PDEs in Table 1 computed from 200 instantiations of noise for each noise level σ_{NR}. Figure from Messenger and Bortz (2021a).

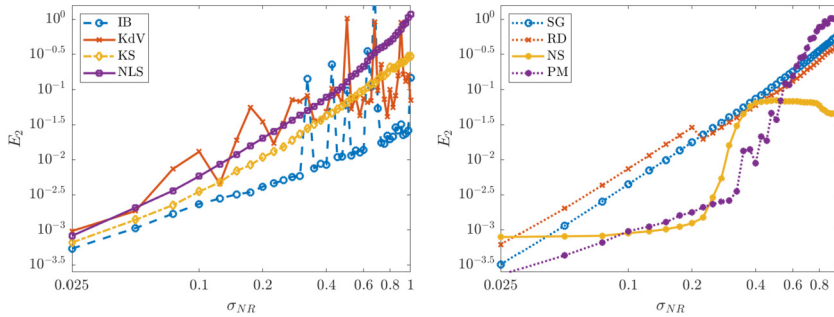


FIGURE 4 Coefficient errors E_2 for each of the seven models Table 1. The plot on the left contains PDEs in one spatial dimension, and the plot on the right contains PDEs in two dimensions. Figure from Messenger and Bortz (2021a).

3 Theoretical results

In the application of WSINDy to several classes of differential equations, the general pattern has emerged that the system identification using the weak form offers substantial advantages over using the strong form. However, there is no rigorous explanation for this behavior and thus a theoretical analysis of WSINDy is an ongoing project. Here we report briefly on recent theoretical results presented in Messenger and Bortz (2022a) concerning the performance of WSINDy in the limit of continuum data. Theorem 1 below (Messenger and Bortz, 2022a, Theorem 4.2) demonstrates that within a broad class of models, the WSINDy estimate converges in probability to the correct model, provided that the noise level σ is below a critical noise threshold σ_c (see Assumptions A for a full list of assumptions). This provides the following:

(I) An explanation for the empirically observed robustness of weak-form equation learning methods
(II) Quantification of the effects of nonsmoothness (e.g. weak solutions) on the rate of convergence
(III) Specification of a class of models for which WSINDy converges for any noise level (i.e. $\sigma_c = \infty$)

In addition, in Messenger and Bortz (2022a, Theorem 4.3) it is also proved that suitably denoising the data (e.g. with a simple moving-average filter) results in unconditional convergence of WSINDy over the class of models with locally-Lipschitz nonlinearities, extending the class specified in (III) above.

3.1 Assumptions

Theorem 1 concerns the performance of WSINDy in the limit of *continuum data*. Using the notation from Section 2.4, this limit is defined by a sequence of noisy samples $\{\mathbf{U}^{(n)}\}_{n=1}^{\infty}$ of the solution u observed on successively finer computational grids $\{(\mathbf{X}^{(n)}, \mathbf{t}^{(n)})\}_{n=1}^{\infty} \subset \Omega \times (0, T)$, each of which is equally spaced with resolution $\Delta x^{(n)}$ in all spatial coordinates and $\Delta t^{(n)}$ in time. For each $\mathbf{U}^{(n)}$, a weak-form linear system $(\mathbf{G}^{(n)}, \mathbf{b}^{(n)})$ is constructed according to (9), and WSINDy is employed to produce an estimate $\widehat{\mathbf{w}}^{(n)}$ of the true weight vector $\mathbf{w}^\star$. We now describe the core assumptions used in the proof of Theorem 1 are below.

We define the following admissible solution spaces for the underlying solution u. For an open, bounded domain $D \subset \mathbb{R}^{d+1}$, codomain $D' \subset \mathbb{R}^N$, $p \in [1, \infty]$, and $k > 0$, define the function spaces

$$\mathcal{H}^{k,p}(D, D') := \Bigg\{ f \in L^p(D, D') \ : \ \exists \text{ disjoint, open } (D_i)_{i=1}^{\ell} \\ \text{s.t. } \overline{D} = \bigcup_{i=1}^{\ell} \overline{D}_i \,, \ f\big|_{D_i} \in H^k(D_i, D'), \ \partial D_i \in C^{0,1} \Bigg\}, \tag{13}$$

where $H^k(D, D')$ is the space of functions from D to D' with weak derivatives up to order k in $L^2(D, D')$. In what follows, we take $D' = \mathbb{R}^N$ and suppress explicit reference to the codomain. The spaces $\mathcal{H}^{k,p}(D) \subset L^p(D)$ are similar to the broken Sobolev spaces used in the analysis of discontinuous Galerkin methods (see e.g. Houston et al., 2002). With $u \in \mathcal{H}^{k,\infty}(\Omega \times (0, T))$ such that $k > (d+1)/2$, we have that pointwise evaluations of u are well-defined (apart from a set of measure zero, e.g. when considering solutions with shocks) by the Sobolev embedding theorem.

In addition, define the *discretization level* m_n by[5]

$$m_n := \#\{\text{supp}\,(\psi) \cap (\mathbf{X}^{(n)}, \mathbf{t}^{(n)})\}. \tag{14}$$

[5] $\#\{\cdot\}$ indicates the set cardinality.

In words, m_n is the number of points that the reference test function ψ is supported on within the grid $(\mathbf{X}^{(n)}, \mathbf{t}^{(n)})$. In Messenger and Bortz (2022a, Theorem 3.1), it is proved that $(\mathbf{G}^{(n)}, \mathbf{b}^{(n)})$ concentrates around its mean at a rate $\mathcal{O}(\exp(-cm_n^{2/p}))$, where c is a universal constant and $p_{\max}$ is the maximum polynomial degree in the model library.

Assumptions A. Let $p \geq 1$ be fixed.

(A.1) (Regularity of u) $u \in \mathcal{H}^{k,\infty}(\Omega \times (0,T))$ for some $k > (d+1)/2$ is a weak solution to (7) with coefficients $\mathbf{w} = \mathbf{w}^\star$.

(A.2) (Noise distribution) Measurement noise $\epsilon = \mathbf{U}^{(n)} - u(\mathbf{X}^{(n)}, \mathbf{t}^{(n)})$ is i.i.d. according to a symmetric and sub-Gaussian probability distribution ρ.[6] We refer to the standard deviation $\sigma = \sqrt{\mathbb{V}[\rho]}$ as the *noise level*.

(A.3) (Computational grids) Each grid $(\mathbf{X}^{(n)}, \mathbf{t}^{(n)})$ has uniform spacing $\Delta x^{(n)}$ in all spatial dimensions and $\Delta t^{(n)}$ in time, and the collection of grids is dense in $\Omega \times (0,T)$, or $\overline{\bigcup_{n=1}^{\infty}(\mathbf{X}^{(n)}, \mathbf{t}^{(n)})} = \Omega \times (0,T)$.

(A.4) (Model library) The family of functions $\mathcal{F} = (f_j(u))_{j\in[J]}$ consists of $P^{(p_{\max})}$, the set of monomials of total degree[7] at most $p_{\max}$ on $\mathbb{R}^N$, as well as $F^{\boldsymbol{\omega}} = \{\exp(i\omega^T u)\}_{\omega\in\boldsymbol{\omega}}$, a finite collection of Fourier modes on $\mathbb{R}^N$ (i.e. $\boldsymbol{\omega} \subset \mathbb{R}^N$ is a finite set). Furthermore we assume[8] $\widehat{\rho}(\omega) \neq 0$ for $\omega \in \boldsymbol{\omega}$.

(A.5) (Reference test function and Query points) We assume that $\psi \in C^{|\boldsymbol{\alpha}|}(\Omega \times (0,T))$ with compact support in $\Omega \times (0,T)$, and that for all $(\mathbf{x}_k, t_k) \in \mathcal{Q}$,

$$\operatorname{supp}(\psi(\mathbf{x}_k - \cdot, t_k - \cdot)) \subset \Omega \times (0,T) \tag{15}$$

(A.6) (Conditioning of $(\mathbf{G}^\star, \mathbf{b}^\star)$) The noise-free continuum matrix $\mathbf{G}^\star$[9] has full column rank. Moreover, the true dynamics have a stable representation in weak-form quantified by

$$\mu^\star := \min_{S \subsetneq S^\star} \frac{\left\| \mathbf{P}^{\perp}_{\mathbf{G}^\star_{S^\star\setminus S}} \mathbf{b}^\star \right\|}{\|\mathbf{b}^\star\|} - \frac{|S|+1}{\mathfrak{J}} > 0 \tag{16}$$

where $\mathfrak{J}$ is number of columns in $\mathbf{G}^\star$, $S^\star$ is the support of the true weight vector $\mathbf{w}^\star$, and $\mathbf{P}^{\perp}_{\mathbf{G}^\star_{S^\star\setminus S}}$ denotes the projection onto space orthogonal to the columns $\mathbf{G}^\star_{S^\star\setminus S}$ lying in the set $S^\star \setminus S$. In words, $\mathbf{b}^\star$ cannot be approximated arbitrarily well from a strict subset of $\mathbf{G}^\star_{S^\star}$.

Under these assumptions, we have the following.

---

[6] That is, ρ satisfies $\|\rho\|_{\mathrm{SG}} := \inf\{\lambda > 0 : \mathbb{E}_{\epsilon\sim\rho}\left[\exp(\epsilon^2/\lambda^2)\right] \leq 2\} < \infty$. This includes e.g. Gaussian and uniform white noise, see Vershynin (2018) for more details.

[7] For a monomial $p(x_1, \ldots, x_N) = \prod_{i=1}^N x_i^{q_i}$, the total degree is defined by $\sum_{i=1}^N q_i$.

[8] $\widehat{\rho}(\omega) = \int_{\mathbb{R}^N} e^{i\omega\cdot y}\rho(y)dy$ is the Fourier transform of ρ.

[9] Entries $\mathbf{G}^\star_{k,(s-1)J+j}$ and $\mathbf{b}^\star_k$ are given by the right and left sides of (9).

Theorem 1. *Provided Assumptions A hold, there exists a critical noise level* $\sigma_c > 0$ *and a stability tolerance* θ_* *such that for all* $\sigma < \sigma_c$, $\theta < \theta_*$, *and sufficiently large n, it holds that*

$$supp\left(\widehat{\mathbf{w}}^{(n)}\right) = supp\left(\mathbf{w}^\star\right) \qquad and \qquad \left\|\widehat{\mathbf{w}}^{(n)} - \mathbf{w}^\star\right\|_\infty < C'(\theta + \sigma^2) \tag{17}$$

with probability exceeding $1 - 4K(\mathfrak{J}+1)\exp\left(-\frac{c}{2}(m_n\theta)^{2/p_{\max}}\right)$, *where* $\widehat{\mathbf{w}}^{(n)} = \text{MSTLS}^{(1)}(\mathbf{G}^{(n)}, \mathbf{b}^{(n)})$ *and* $c, C' > 0$ *are independent of n.*

The proof of Theorem 1 is contained in Messenger and Bortz (2022a, Theorem 4.2).[10]

4 Weak form-based parameter estimation

The previous sections describe how WSINDy can be used to discover governing equations. However, in the case where there is high confidence in the model equation itself, one can also perform parameter estimation using the weak form of the model. Versions of this idea have existed since the mid 1950's (Shinbrot, 1954), but the proposed test functions either did not have enough smoothness or were spectrally mismatched with the data and did not yield highly accurate estimates. Accordingly, a natural research direction is to build on the success of WSINDy and combine it with the theoretical results in §3 as well as modern statistical regression to create a novel parameter estimation method. In Bortz et al. (2023), we introduced an improved weak-form parameter estimation algorithm WENDy (Weak-form Estimation of Nonlinear Dynamics) which works for the class of differential equations with right sides that are linearly separable. In §4.1, we will summarize the results for ODEs and then discuss how to apply WENDy to PDEs §4.2 and SDEs §4.3.

In this section we assume that the model is known and will attempt to solve this parameter estimation problem

$$\widehat{\mathbf{w}} := \arg\min_{\mathbf{w}\in\mathbb{R}^J} \|u(\mathbf{t};\mathbf{w}) - \mathbf{U}\|_2^2, \tag{18}$$

where the data $\mathbf{U} \in \mathbb{R}^{(M+1)\times d}$ is sampled at $M+1$ timepoints $t := \{t_i\}_{i=0}^M$, and the function $u : \mathbb{R} \to \mathbb{R}^d$ is a solution to a differential equation model

$$\begin{aligned} \dot{u} &= \sum_{j=1}^J w_j f_j(u), \\ u(t_0) &= u_0 \in \mathbb{R}^d. \end{aligned} \tag{19}$$

[10] Also note that MSTLS$^{(1)}$ refers to the Modified Sequential Thresholding Least Squares algorithm with a single thresholding step per $\lambda \in \boldsymbol{\lambda}$, or $\ell = 0$ (see Messenger and Bortz, 2021a).

4.1 Ordinary differential equations

We begin by considering a d-dimensional matrix form of (19), i.e., an ordinary differential equation system model

$$\dot{u} = \Theta(u)W \tag{20}$$

with row vector of the d solution states

$$u(t; W) := [\ u_1(t; W) \ \big|\ u_2(t; W) \ \big|\ \cdots \ \big|\ u_d(t; W)]\ ,$$

row vector of J features (i.e., right side terms where $f_j : \mathbb{R}^d \to \mathbb{R}$ is C_c^2) such that $\Theta(u) := [\ f_1(u) \ \big|\ f_2(u) \ \big|\ \cdots \ \big|\ f_J(u)]$, and the matrix of unknown parameters $W \in \mathbb{R}^{J\times d}$. The matrix version of these terms evaluated at the time-points is thus

$$\mathbf{t} := \begin{bmatrix} t_0 \\ \vdots \\ t_M \end{bmatrix}, \qquad \mathbf{u} := \begin{bmatrix} u_1(t_0) & \cdots & u_d(t_0) \\ \vdots & \ddots & \vdots \\ u_1(t_M) & \cdots & u_d(t_M) \end{bmatrix},$$

$$\Theta(\mathbf{u}) := \begin{bmatrix} f_1(u(t_0)) & \cdots & f_J(u(t_0)) \\ \vdots & \ddots & \vdots \\ f_1(u(t_M)) & \cdots & f_J(u(t_M)) \end{bmatrix}.$$

Multiplication by a set of compactly supported test functions $\{\phi_{m_k}\}$ (centered at timepoints $\{t_{m_k}\}$, a subset of the entries of $\mathbf{t}$) using trapezoidal quadrature yields

$$-\dot{\Phi}_k \mathbf{u} \approx \Phi_k \Theta(\mathbf{u}) W, \tag{21}$$

where

$$\Phi_k := \left[\ \phi_k(t_0) \ \big|\ \cdots \ \big|\ \phi_k(t_M) \ \right], \qquad \dot{\Phi}_k := \left[\ \dot{\phi}_k(t_0) \ \big|\ \cdots \ \big|\ \dot{\phi}_k(t_M) \ \right].$$

The core idea of the weak-form-based direct parameter estimation is to identify W as a least squares solution to

$$\min_W \|\mathrm{vec}(\mathbf{G}W - \mathbf{B})\|_2^2 \tag{22}$$

where "vec" vectorizes a matrix,

$$\mathbf{G} := \Phi\Theta(\mathbf{U}) \in \mathbb{R}^{K\times J},$$
$$\mathbf{B} := -\dot{\Phi}\mathbf{U} \in \mathbb{R}^{K\times d},$$

where $\mathbf{U} \in \mathbb{R}^{(M+1)\times d}$ represents the data, and the integration matrices are

$$\Phi = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_K \end{bmatrix} \in \mathbb{R}^{K\times(M+1)} \quad \text{and} \quad \dot{\Phi} = \begin{bmatrix} \dot{\Phi}_1 \\ \vdots \\ \dot{\Phi}_K \end{bmatrix} \in \mathbb{R}^{K\times(M+1)}.$$

4.1.1 WENDy using iterative reweighting

We note that the posed regression problem does not fit within the framework of ordinary least squares, and is actually an Errors-In-Variables problem. We will also derive a linearization that reveals a covariance structure which depends on the jacobian of the right side as well as the true parameters. First, we denote the vector of true (but unknown) parameter values used in all state variable equations as $\mathbf{w}^\star$ and let $u^\star := u(t; \mathbf{w}^\star)$ and $\Theta^\star := \Theta(u^\star)$. The system measurements are assumed to be noisy so that at each timepoint t all states are observed with additive noise

$$U(t) = u^\star(t) + \varepsilon(t) \tag{23}$$

where each element of $\varepsilon(t)$ is i.i.d. $\mathcal{N}(0, \sigma^2)$. Lastly, we note that there are d variables, J feature terms, and $M+1$ timepoints. In what follows, we present the expansion using Kronecker products (denoted as $\otimes$).

We begin by considering the sampled data $\mathbf{U} := \mathbf{u}^\star + \boldsymbol{\varepsilon} \in \mathbb{R}^{(M+1)\times d}$ and vector of parameters to be identified $\mathbf{w} \in \mathbb{R}^{Jd}$. We use bolded variables to represent evaluation at the timegrid $\mathbf{t}$, and use superscript $\star$ notation to denote quantities based on true (noise-free) parameter or states. We now consider the residual

$$\mathbf{r}(\mathbf{U}, \mathbf{w}) := \mathbf{G}\mathbf{w} - \mathbf{b}, \tag{24}$$

where we redefine

$$\begin{aligned} \mathbf{G} &:= [\mathbb{I}_d \otimes (\Phi\Theta(\mathbf{U}))], \\ \mathbf{b} &:= -\text{vec}(\dot{\Phi}\mathbf{U}). \end{aligned}$$

We decompose and linearize $\mathbf{r}$ such that

$$\mathbf{r}(\mathbf{U}, \mathbf{w}) \approx \mathbf{G}^\star\mathbf{W} - \mathbf{b}^\star + \mathbf{L}_\mathbf{w}\text{vec}(\boldsymbol{\varepsilon}), \tag{25}$$

where

$$\begin{aligned} \mathbf{G}^\star &:= [\mathbb{I}_d \otimes (\Phi\Theta(\mathbf{u}^\star))], \\ \mathbf{b}^\star &:= -\text{vec}(\dot{\Phi}\mathbf{u}^\star), \\ \mathbf{L}_\mathbf{w} &:= [\text{mat}(\mathbf{w})^T \otimes \Phi]\nabla\Theta\mathbf{K} + [\mathbb{I}_d \otimes \dot{\Phi}], \end{aligned}$$

where "mat" is the matricization operation and $\mathbf{K}$ is the commutation matrix such that $\mathbf{K}\text{vec}(\boldsymbol{\varepsilon}) = \text{vec}(\boldsymbol{\varepsilon}^T)$. The matrix $\nabla\Theta$ contains derivatives of the features

$$\nabla\Theta := \left[\begin{array}{ccc} \nabla f_1(\mathbf{U}_0) & & \\ & \ddots & \\ & & \nabla f_1(\mathbf{U}_M) \\ \hline & \vdots & \\ \hline \nabla f_J(\mathbf{U}_0) & & \\ & \ddots & \\ & & \nabla f_J(\mathbf{U}_M) \end{array}\right], \tag{26}$$

where

$$\nabla f_j(\mathbf{U}_m) = \left[\begin{array}{c|c|c} \frac{\partial}{\partial u_1} f_j(\mathbf{U}_m) & \cdots & \frac{\partial}{\partial u_d} f_j(\mathbf{U}_m) \end{array}\right],$$

and $\mathbf{U}_m \in \mathbb{R}^{1\times d}$ is the row vector of data at t_m.

If all elements of $\boldsymbol{\varepsilon}$ are i.i.d. Gaussian, i.e., $\mathcal{N}(0, \sigma^2)$ then to first order

$$\mathbf{r}(\mathbf{U}, \mathbf{w}) - (\mathbf{r}_0 + \mathbf{e}_{\text{int}}) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{L}_{\mathbf{w}} (\mathbf{L}_{\mathbf{w}})^T). \tag{27}$$

We note that in (27), the covariance is dependent upon the parameter vector $\mathbf{w}$. In the statistical inference literature, the Iteratively Reweighted Least Squares (IRLS) (Jorgensen, 2012) method provide a good strategy to account for a parameter-dependent covariance by iterating between solving for $\mathbf{w}$ and updating the covariance matrix $\mathbf{C}$. In Algorithm 2 we present the WENDy method, updating $\mathbf{C}^{(n)}$ (at the n-th iteration step) in lines 7-8 and then the new parameters $\mathbf{w}^{(n+1)}$ are computed in line 9 by weighted least squares.

Depicted in Fig. 5 are the results of comparing WENDy and FSNLS for the FitzHugh-Nagumo equation on a scatterplot of walltime (in seconds) vs. the relative accuracy in the estimated parameters.

$$\begin{aligned} \dot{u}_1 &= 3u_1 - 3u_1^3 + 3u_2 \\ \dot{u}_2 &= -\tfrac{1}{3}u_1 + \tfrac{17}{150} + \tfrac{1}{15}u_2 \end{aligned}$$

It is clear in this figure that WENDy is on average both more accurate and faster. In Bortz et al. (2023), WENDy is applied to several other benchmark ODE parameter estimation problems including Logistic growth, Lotka-Volterra, Hindmarsh-Rose, and a Protein Transduction system. In almost all cases, WENDy offered more accurate parameter estimates. And, in *all* cases, WENDy was faster, sometimes by several orders of magnitude.

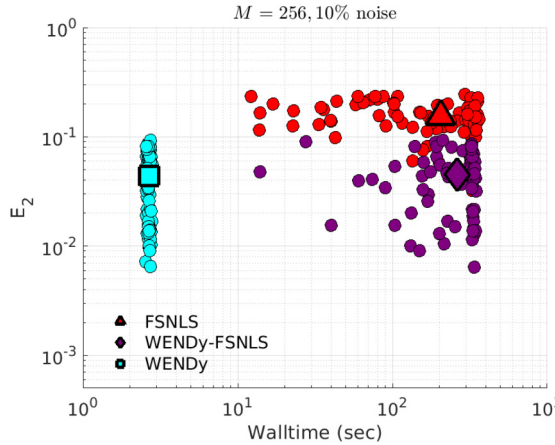


FIGURE 5 Comparison between Forward Solver-based Nonlinear Least Squares (FSNLS), Weak form Estimation of Nonlinear Dynamics (WENDy), and using the results of WENDy as an initial estimate for FSNLS (WENDy-FSNLS) for the FitzHugh-Nagumo model. Both variables are observed at 256 timepoints and with 10% additive Gaussian noise level. Figure from Bortz et al. (2023).

4.2 Partial differential equations

Extending WENDy to PDEs is straightforward. Consider the weak form (9) written as a convolution between the test function and the equation terms, and assume that the ith solution component obeys

$$\left(\mathscr{D}^{\alpha^0}\psi\right) * u_i(\mathcal{Q}) - \sum_{j=1}^{J} \mathbf{W}_{ji}\left(\mathscr{D}^{\alpha^j}\psi\right) * f_j(u)(\mathcal{Q}) = 0.$$

We can write this in discrete form as

$$\Phi^{\alpha_0}\mathbf{U}_i^\star - \sum_{j=1}^{J} \Phi^{\alpha_j} f_j(\mathbf{U}^\star)\mathbf{W}_i = \mathbf{b}_i - \mathbf{G}\mathbf{W}_i \approx 0$$

where Φ^{α^j} are associated discrete matrices that enact the convolution with $\mathscr{D}^{\alpha^j}\psi$ over spacetime, evaluated at the query points $\mathcal{Q}$, and $\mathbf{W}_i$ is the ith column of $\mathbf{W}$. Performing similar analysis to Section 4.1.1, we arrive at the linear transformation

$$\mathbf{L}_{\mathbf{w}} := \left[\mathbf{W}_1^T \otimes \Phi^{\alpha_1} \quad \mathbf{W}_2^T \otimes \Phi^{\alpha_2} \quad \cdots \quad \mathbf{W}_J^T \otimes \Phi^{\alpha_J}\right]\nabla\Theta\mathbf{K} + \left[\mathbb{I}_d \otimes \Phi^{\alpha^0}\right].$$

Here $\mathbf{W}_j^T$ is the jth column of $\mathbf{W}$ and $\nabla\Theta$ is as in (26) suitably reindexed to account for the vectorization of multidimensional arrays. Appealing to sparse matrix constructions, the results can similarly accelerate and improve on estimates obtained using forward simulation-based nonlinear least squares, as

Algorithm 2: WENDy

input : Data $\{\mathbf{U}\}$, Feature Map $\{\Theta, \nabla\Theta\}$, Test Function Matrices $\{\Phi, \dot{\Phi}\}$, Stopping Criteria {SC}, Covariance Relaxation Parameter $\{\alpha\}$, Variance Filter $\{\mathbf{f}\}$

output: Parameter Estimate $\{\widehat{\mathbf{w}}, \widehat{\mathbf{C}}, \widehat{\sigma}, \mathbf{S}, \texttt{stdx}\}$

```
// Compute weak-form linear system
```
$\mathbf{G} \leftarrow [\mathbb{I}_d \otimes (\Phi\Theta(\mathbf{U}))]$

$\mathbf{b} \leftarrow -\text{vec}(\dot{\Phi}\mathbf{U})$

```
// Solve Ordinary Least Squares Problem
```
$\mathbf{w}^{(0)} \leftarrow (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{b}$

```
// Solve Iteratively Reweighted Least Squares Problem
```
$n \leftarrow 0$

`check` $\leftarrow$ true

while `check` is true **do**

  $\mathbf{L}^{(n)} \leftarrow [\text{mat}(\mathbf{w}^{(n)})^T \otimes \Phi]\nabla\Theta(\mathbf{U})\mathbf{K} + [\mathbb{I}_d \otimes \dot{\Phi}]$

  $\mathbf{C}^{(n)} = (1-\alpha)\mathbf{L}^{(n)}(\mathbf{L}^{(n)})^T + \alpha\mathbf{I}$

  $\mathbf{w}^{(n+1)} \leftarrow (\mathbf{G}^T(\mathbf{C}^{(n)})^{-1}\mathbf{G})^{-1}\mathbf{G}^T(\mathbf{C}^{(n)})^{-1}\mathbf{b}$

  `check` $\leftarrow \text{SC}(\mathbf{w}^{(n+1)}, \mathbf{w}^{(n)})$

  $n \leftarrow n+1$

end

```
// Return estimate and standard statistical quantities
```
$\widehat{\mathbf{w}} \leftarrow \mathbf{w}^{(n)}$

$\widehat{\mathbf{C}} \leftarrow \mathbf{C}^{(n)}$

$\widehat{\sigma} \leftarrow (Md)^{-1/2}\|\mathbf{f} * \mathbf{U}\|_{\text{F}}$

$\mathbf{S} \leftarrow \widehat{\sigma}^2((\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T)\,\widehat{\mathbf{C}}\,(\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}))$

`stdx` $\leftarrow \sqrt{\texttt{diag}(\mathbf{S})}$

evidenced in Fig. 6 for the Kuramoto-Sivashinsky PDE. All circles and squares in the figure represent different realizations of the noisy data. While there is certainly some variability, on average WENDy is two orders of magnitude faster and an order of magnitude more accurate than the conventional FSNLS method.

4.3 Stochastic differential equations

Weak-form estimation can also be easily extended to stochastic differential equations. The focus is to identify the drift f and diffusivity σ of an Itô SDE

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t \tag{28}$$

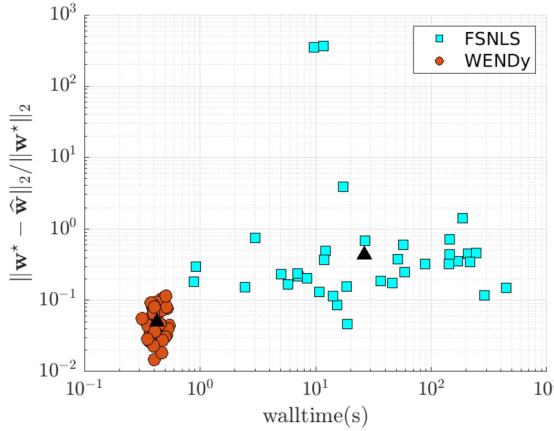


FIGURE 6 Coefficient error vs. walltime for WENDy applied to the Kuramoto-Sivashinsky equation, as described in Section 4.2, vs. forward solver-based nonlinear least-squares (FSNLS). The underlying data has 128×64 points in space and time and 20% added Gaussian white noise.

given discrete-time observations of its solution $X_t \in \mathbb{R}^d$. In this setting the input data is a collection of L discrete-time realizations $\mathbf{Y} = \{\mathbf{Y}^{(\ell)}\}_{\ell=1}^{L}$, where each realization $\mathbf{Y}^{(\ell)} = (Y_{t_0}^{(\ell)}, \ldots, Y_{t_{M+1}}^{(\ell)})$ occurs over the time grid $\mathbf{t} = (t_0, \ldots, t_{M+1})$, and at each time $t_i \in \mathbf{t}$ the observations are given by $Y_{t_i}^{(\ell)} = X_{t_i}^{(\ell)} + \epsilon$. Here ϵ represents possible measurement noise and $X_t^{(\ell)}$, $\ell = 1, \ldots, L$ are solutions to (28) with initial conditions $x_0^{(\ell)}$ each drawn independently from the distribution ρ_0.

Weak formulation in time

Itô calculus can be employed to formulate the weak-form discovery problem. Let $\psi(x,t) : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ be a C^2 function compactly-supported in the time interval $(0,T)$ for all $x \in \mathbb{R}^d$. Itô's formula applied (28) then gives

$$d\left(\psi(X_t,t)\right) = \left(\partial_t\psi(X_t,t) + \nabla\psi(X_t,t)\cdot f(X_t,t) + \frac{1}{2}\mathrm{H}\psi(X_t,t) : (\sigma\sigma^T)\right)dt + \nabla\psi(X_t,t)\cdot\sigma(X_t,t)dB_t,$$

where H denotes the Hessian and $A : B = \mathrm{vec}(A)\cdot\mathrm{vec}(B)$. Integrating in time and using compact support, we get

$$-\int_0^T \partial_t\psi(X_t,t)\,dt = \int_0^T \left(\nabla\psi(X_t,t)\cdot f(X_t,t) + \frac{1}{2}\mathrm{H}\psi(X_t,t) : (\sigma\sigma^T)\right)dt + \int_0^T \nabla\psi(X_t,t)\cdot\sigma(X_t,t)dB_t. \tag{29}$$

Estimation algorithm: test and trial functions

As in the PDE case, the set of test functions $\Psi = \{\psi_1, \ldots, \psi_K\}$ can be chosen in a flexible and efficient manner by letting each ψ_k be separable, of the form

$$\psi_k(x,t) = \theta_k(t)\phi_k(x).$$

This leads to a very general scheme, with the only requirements being that $\theta_k \in C^1$, compactly support in $[0,T]$, and $\phi_k \in C^2$.

We search for linear representations of the drift f and squared diffusivity $\sigma\sigma^T$. That is, we define a drift basis $\mathbb{F} = (f_1, \ldots, f_{J_f})$ and a basis of upper-triangular matrices $\mathbb{S} = (\Sigma_1, \ldots, \Sigma_{J_\sigma})$, and search for linear representations

$$\widehat{f} = \sum_{j=1}^{J_f} w_j^{\mathbb{F}} f_j, \qquad \widehat{\Sigma} = \sum_{j=1}^{J_\sigma} w_j^{\mathbb{S}} \Sigma_j.$$

Our diffusivity estimator is then

$$\widehat{\sigma\sigma^T} = \widehat{\Sigma} + \widehat{\Sigma}^T - \operatorname{diag}(\widehat{\Sigma}).$$

Note that for two symmetric matrices A, B, letting $U(A)$ and $U(B)$ denote their upper-triangular parts, we have

$$\begin{aligned} A : B &= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij} = 2\sum_{i=1}^{n}\sum_{j=i}^{n} A_{ij}B_{ij} - \operatorname{diag}(A) : \operatorname{diag}(B) \\ &=: (U(A) : U(B)) \end{aligned}$$

which defines an inner product $(\cdot : \cdot)$ on upper triangular matrices, so the representation of the squared diffusivity $\widehat{\sigma\sigma^T}$ above in terms of $\mathbb{S}$ is valid.

Quadrature

Integrals in (28) are discretized using the trapezoidal rule, which achieves an optimal minimax rate for integrands with the same regularity as Brownian motion (Diaconis, 1988), which is shared by solutions to Itô SDEs.[11] For each observed realization $\mathbf{Y}^{(\ell)}$, we define the matrices $(\mathbf{G}^{\mathbb{F}}(\mathbf{Y}^{(\ell)}), \mathbf{G}^{\mathbb{S}}(\mathbf{Y}^{(\ell)}))$ by

$$\mathbf{G}^{\mathbb{F}}_{kj}(\mathbf{Y}^{(\ell)}) = \sum_{i=1}^{M} \nabla\psi_k(\mathbf{Y}^{(\ell)}_{t_i}, t_i) \cdot f_j(\mathbf{Y}^{(\ell)}_{t_i}, t_i) \left(\frac{\Delta t_i + \Delta t_{i-1}}{2}\right)$$

$$\mathbf{G}^{\mathbb{S}}_{kj}(\mathbf{Y}^{(\ell)}) = \sum_{i=1}^{M} \left(\mathrm{H}\psi_k(\mathbf{Y}^{(\ell)}_{t_i}, t_i) : \Sigma_j(\mathbf{Y}^{(\ell)}_{t_i}, t_i)\right) \left(\frac{\Delta t_i + \Delta t_{i-1}}{2}\right),$$

[11] Strictly speaking, if the diffusivity is nonconstant, then regularity can be slightly worse than this, in which case there may be a better quadrature. It may even be the case that Riemann sums have lower variance, this needs to be fully explored.

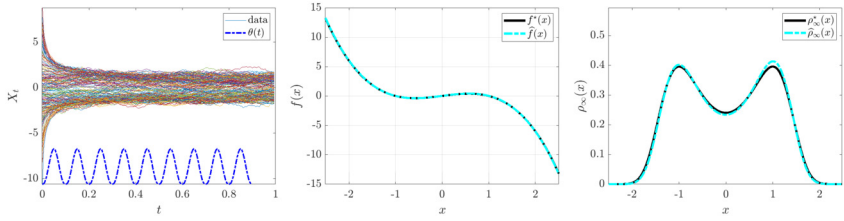


FIGURE 7 WENDy applied to the stochastic double-well potential dynamics (30). Left: data realizations and temporal testfunctions ϕ_k. Middle: comparison between learned ($\hat{f}$) and true ($f^\star$) drift functions. Right: comparison between learned ($\hat{\rho}_\infty$) and true ($\rho^\star_\infty$) stationary measure.

where timesteps are given by $\Delta t_i = t_{i+1} - t_i$. We then form the concatenated matrix

$$\mathbf{G}(\mathbf{Y}^{(\ell)}) = \left[\mathbf{G}^{\mathbb{F}}(\mathbf{Y}^{(\ell)}) \quad \mathbf{G}^{\mathbb{S}}(\mathbf{Y}^{(\ell)})\right].$$

The response vector $\mathbf{b}(\mathbf{Y}^{(\ell)})$ is defined by

$$\mathbf{b}_k(\mathbf{Y}^{(\ell)}) = -\sum_{i=1}^{M} \partial_t \psi_k(\mathbf{Y}^{(\ell)}_{t_i}, t_i)\left(\frac{\Delta t_i + \Delta t_{i-1}}{2}\right).$$

It then holds that

$$\mathbf{b}(\mathbf{Y}^{(\ell)}) \approx \mathbf{G}(\mathbf{Y}^{(\ell)})\mathbf{w}$$

where $\mathbf{w} = ((\mathbf{w}^{\mathbb{F}})^T, (\mathbf{w}^{\mathbb{S}})^T)^T$. To combine the L observed realizations $\mathbf{Y}^{(\ell)}$, we average the L linear systems together, arriving at a small linear system $(\overline{\mathbf{G}}, \overline{\mathbf{b}})$, which reduces variance at the Monte-Carlo rate in L.

Optimization

The residuals

$$\mathbf{r}^{(\ell)}(\mathbf{w}) = \mathbf{b}(\mathbf{Y}^{(\ell)}) - \mathbf{G}(\mathbf{Y}^{(\ell)})\mathbf{w}$$

are amenable to the same analysis employed in WENDy in the context of ODEs and PDEs, leading to an iteratively reweighted least-squares approach for the weights $\widehat{\mathbf{w}}$. Specifically, the right-most term in (29) is an Itô integral and constitutes the residual for each test function. Its covariance against other residuals can be computed using Itô calculus and properties of local martingales. However, as a first pass ordinary least squares can be used, which can be seen to perform well applied to the double-well potential dynamics (30) (see Fig. 7).

Test problem: double-well potential

Consider $X_t \in \mathbb{R}$ satisfying (28)

$$f(x,t) = x - x^3, \qquad \sigma(x,t) = 1. \tag{30}$$

The dynamics, drift and stationary measure are pictured in Fig. 7.

TABLE 3 Comparison between state-of-the-art deep learning-based estimation and WENDy.

Parameter	λ_0	λ_1	λ_2	λ_3	μ	runtime
True value	0	1	0	−1	1	—
PINN (Chen et al., 2021)	0.0051	0.8422	−0.0071	−0.8994	1.0347	minutes-hours
WENDy	0.0064	1.0145	0.00045	−1.0071	0.958	2.5 sec

An existing PINN-based approach by Chen et al. (2021) for learning f and σ in (30) used $\mathbf{Y}$ containing 40,000 datapoints $Y_{t_i}^{(\ell)}$. Runtime information is not available in Chen et al. (2021), however can reasonably be estimated from the method. The authors used an architecture of 4 hidden layers each with 20 neurons and performed optimization via Adam with learning rate 10^{-4} and 200,000 iterations. The high iteration count and slow learning rate, combined with optimization over an 80-dimensional space, is guaranteed to far exceed the $O(1)$ seconds required for WENDy, and so is simply listed as "minutes-hours" in Table 3.

To compare with WENDy, we use the same amount of data spread over $L = 200$ realizations $\mathbf{Y}^{(\ell)}$ each containing $M + 2 = 200$ timepoints $t_0 = 0, \ldots, 1 = t_{M+1}$. We use the same cubic trial drift basis, $\mathbb{F} = (1, x, x^2, x^3)$, as in Chen et al. (2021). For test functions we use $\theta_k(t) = \theta(t - t_k)$ and $\phi_k(x) = \phi(x - x_k)$, for query points $Q = \{(x_k, t_k)\}_{k=1}^K$, with base functions $\theta(t) = (1 - (t/r_t)^2)^3$ and $\psi = \exp(-x^2/(2r_x^2))$. We use ~ 1000 equally-spaced query points.[12] We set $(r_t, r_x) = (5\Delta t, \mu/3)$ so that ϕ and ψ are reasonably localized with respect to $\mathbf{Y}$, however, the radii (r_t, r_x) can easily be informed from the dynamics.

The results are in Table 3. Estimated parameters are significantly more accurate than those in Chen et al. (2021) (taken from Table4: Appendix B). Moreover, WENDy takes approximately 2.5 seconds to run.

As well as parameter estimates, it is instructive to measure the point-wise drift error between the learned drift $\hat{f}$ and f, as well as the resulting stationary measure $\hat{\rho}_\infty$ to the true stationary measure ρ_∞, which can be computed using $\hat{f}$ and $\hat{\mu}$:

$$\hat{\rho}_\infty(x) = Z^{-1} \exp\left(-\hat{\mu}^{-1} \int^x \hat{f}(v) dv\right)$$

where Z is computed such that $\|\hat{\rho}_\infty\|_1 = 1$ using suitable quadrature, and $\hat{f}$ can be integrated exactly, being polynomial. For the results in Fig. 7, we have

$$\frac{\|\hat{f} - f\|_2}{\|f\|_2} = 0.00552, \qquad \|\hat{\rho}_\infty - \rho_\infty\|_2 / \|\rho_\infty\| = 0.0274,$$

[12] Note that local, smooth, decaying functions are advantageous, even though not required for ψ. It can be shown that $\psi_k(x) = x^2$ will amplify the stochastic noise in X_t, decreasing quadrature accuracy.

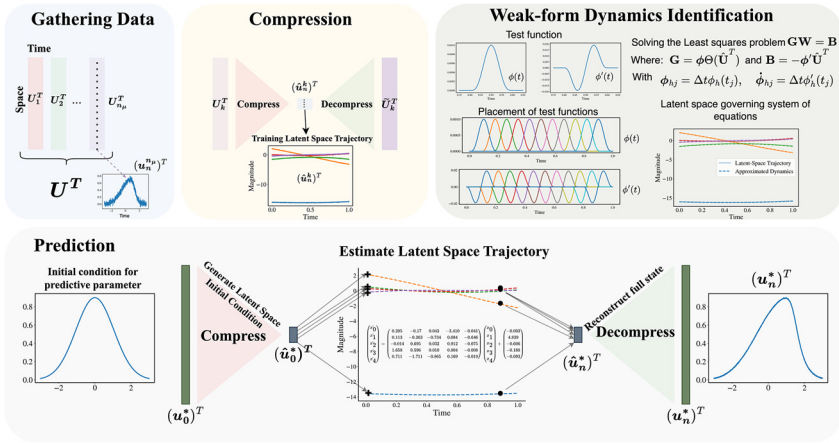


FIGURE 8 Diagram illustrating the WLaSDI algorithm's application to 1D Burgers' simulations, featuring four key steps: data gathering, compression, weak form dynamics identification, and prediction. Diagram from Tran et al. (2024).

providing a highly accurate read on the short-time as well as long-time underlying dynamics.

5 Weak form-based reduced order modeling

Reduced Order Models (ROMs) can be used to accelerate simulations while maintaining high accuracy. The Weak Form Latent Space Dynamics Identification (WLaSDI) employs projection-based reduced order modeling (pROM) in conjunction with weak form parameter estimation. As depicted in Fig. 8, WLaSDI first compresses data, then projects onto the test functions and learns the local latent space models. The variance reduction of the weak form offers robust and precise latent space recovery, hence allowing for a fast and accurate simulation.

To illustrate WLaSDI, consider an N_s-dimensional full-order model, characterized by:

$$\frac{d\mathbf{u}}{dt} = \mathbf{h}(\mathbf{u}, t), \quad \mathbf{u}(0; \boldsymbol{\mu}) = \mathbf{u}_0(\boldsymbol{\mu}) \tag{31}$$

where $t \in [0, T]$. The solution, denoted as $\mathbf{u}(t; \boldsymbol{\mu})$, maps from $[0, T] \times \mathcal{P}$ to $\mathbb{R}^{1 \times N_s}$. The initial condition is given by $\mathbf{u}_0$, where $\boldsymbol{\mu} \in \mathcal{P}$ represents a parameter affecting only the initial conditions. We assume a uniform time step, $\Delta t \in \mathbb{R}$. Throughout this chapter, we use $\mathbf{u}_n := \mathbf{u}(t_n; \boldsymbol{\mu})$.

5.1 WLaSDI algorithm

To begin, WLaSDI draws samples from the parameter space $\mathcal{P}$ and collects together snapshots of artificial data. The sampling points within a training set

$\mathcal{S} \subset \mathcal{P}$ are denoted as $\boldsymbol{\mu}_k$, where $k \in \mathbb{N}(n_\mu)$ and n_μ represents the number of sampling points in the training set. The solution at the n-th time step of Eq. (31) with $\boldsymbol{\mu} = \boldsymbol{\mu}_k$ is represented by $\mathbf{u}_n^k \in \mathbb{R}^{1\times N_s}$ and organized into a snapshot matrix:

$$\mathbf{U}_k = \begin{bmatrix} (\mathbf{u}_0^k)^T & (\mathbf{u}_1^k)^T & \cdots & (\mathbf{u}_{N_t}^k)^T \end{bmatrix}^T \in \mathbb{R}^{(N_t+1)\times N_s}$$

To compile the whole snapshot matrix $\mathbf{U} \in \mathbb{R}^{(N_t+1)n_\mu \times N_s}$, all individual snapshot matrices are concatenated as

$$\begin{bmatrix} \mathbf{U}_0^T & \mathbf{U}_1^T & \cdots & \mathbf{U}_{N_t}^T \end{bmatrix}^T$$

The second step of WLaSDI involves compressing the matrix $\mathbf{U}^T$ using either linear (Proper Orthogonal Decomposition - POD) or nonlinear (Autoencoder) techniques. The linear compression method is referred to as WLaSDI-LS (Linear Subspace), and the nonlinear method is denoted as WLaSDI-NM (Nonlinear Manifold).

- **Proper Orthogonal Decomposition**
  POD generates a spatial basis $\widehat{\boldsymbol{\Psi}}$ that compactly represents $\mathbf{U}^T$, minimizing the projection error:

  $$\widehat{\boldsymbol{\Psi}} := argmin_{\boldsymbol{\Psi}\in\mathbb{R}^{N_s\times n_s}, \boldsymbol{\Psi}^T\boldsymbol{\Psi}=\mathbf{I}} \left\| \mathbf{U}^T - \boldsymbol{\Psi}\boldsymbol{\Psi}^T\mathbf{U}^T \right\|_F^2$$

  We set $\widehat{\boldsymbol{\Psi}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{n_s} \end{bmatrix}$ for $n_s < n_\mu(N_t+1)$. We choose n_s to be the dimension of the latent space. The vector $\mathbf{v}_k$ is the k-th column of the left singular matrix $\mathbf{V}$ from the Singular Value Decomposition (SVD), $\mathbf{U}^T = \mathbf{V}\boldsymbol{\Sigma}\mathbf{W}$. Projecting the snapshot matrix $\mathbf{U}^T$ onto the subspace spanned by the column vectors of $\widehat{\boldsymbol{\Psi}}^T$ results in a reduced snapshot matrix $\widehat{\mathbf{U}}^T \in \mathbb{R}^{n_s\times(N_t+1)n_\mu}$, i.e.,

  $$\widehat{\mathbf{U}}^T := \widehat{\boldsymbol{\Psi}}^T\mathbf{U}^T$$

- **Autoencoder**
  Auto-encoders function as a nonlinear counterpart to POD. Two neural networks undergo training: one for the encoder $\mathcal{G}_{en} : \mathbb{R}^{N_s} \to \mathbb{R}^{n_s}$ and another for the decoder: $\mathcal{G}_{de} : \mathbb{R}^{n_s} \to \mathbb{R}^{N_s}$. The objective is to minimize the mean square error:

  $$MSE(\mathbf{U}^\mathbf{T} - \mathcal{G}_{de}(\mathcal{G}_{en}(\mathbf{U}^\mathbf{T}))$$

  Analogous to POD, we obtain reduced space data through $\widehat{\mathbf{U}}^T := \mathcal{G}_{en}(\mathbf{U}^T) \in \mathbb{R}^{n_s\times(N_t+1)n_\mu}$

With the data projected onto the latent space, we can then use WENDy to construct a surrogate ODE model which is substantially faster to simulate than a

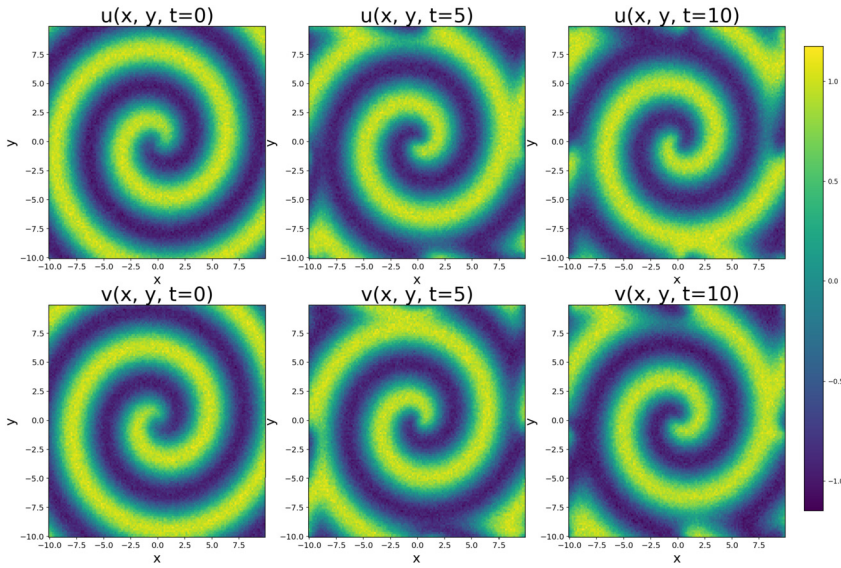


FIGURE 9 Solution to the reaction-diffusion system with 10% noise.

5.2 WLaSDI example

We illustrate WLaSDI using the following reaction-diffusion system:

$$\begin{cases} \partial_t u = \frac{1}{10}\partial_{xx} u + \frac{1}{10}\partial_{yy} u - uv^2 - u^3 + v^3 + u^2 v + u \\ \partial_t v = \frac{1}{10}\partial_{xx} v + \frac{1}{10}\partial_{yy} v + v - uv^2 - u^3 - v^3 - u^2 v \end{cases} \tag{32}$$

The simulation of system (32) takes place across a periodic domain (x, y) within the range of $[-10, 10] \times [-10, 10]$, and the temporal domain spans $[0, 10]$. The simulation employs Fourier spectral differentiation in spatial dimensions and Python's scipy integration. The computational domain is characterized by dimensions $nx = ny = 128$ and $nt = 201$. The initial condition for the system takes the form of a spiral, expressed parametrically as follows:

$$\begin{cases} u(x, y, 0; a, b) = \tanh(a\sqrt{x^2 + y^2}) \cos(\theta - b\sqrt{x^2 + y^2}) \\ v(x, y, 0; a, b) = \tanh(a\sqrt{x^2 + y^2}) \sin(\theta - b\sqrt{x^2 + y^2}) \end{cases}$$

This results in unstable spirals that break apart over time. An example of the solution is provided with parameters $a = b = 1$ and 10% added white noise as shown in Fig. 9.

The parameter space is defined by $a = [0.9, 1.0.1.1]$ and $b = 0.9, 1.0, 1.1$ The testing parameter $\mu^* = (0.95, 1.05)$. In the compression step, we employ Proper Orthogonal Decomposition (POD) on the noisy data, selecting a latent

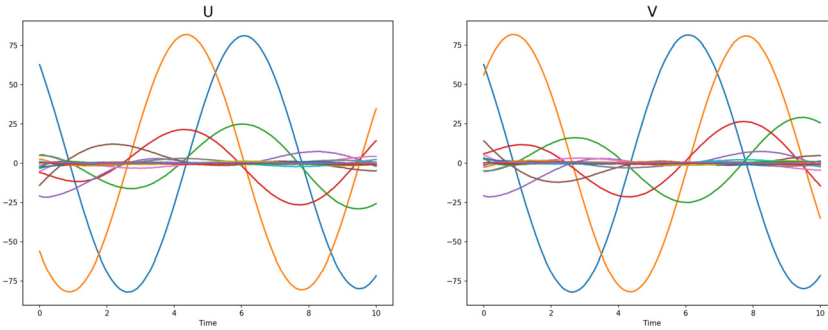
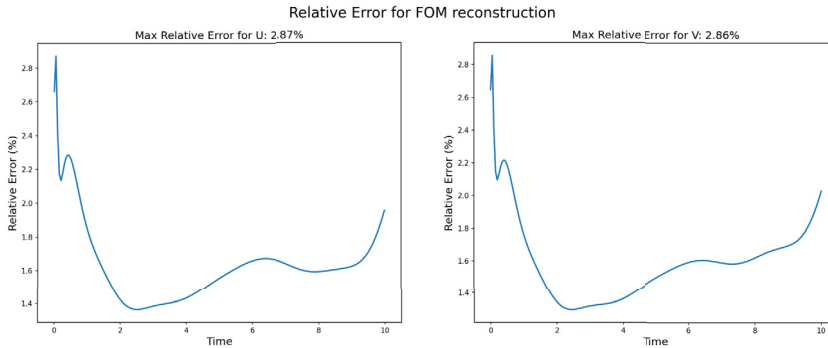
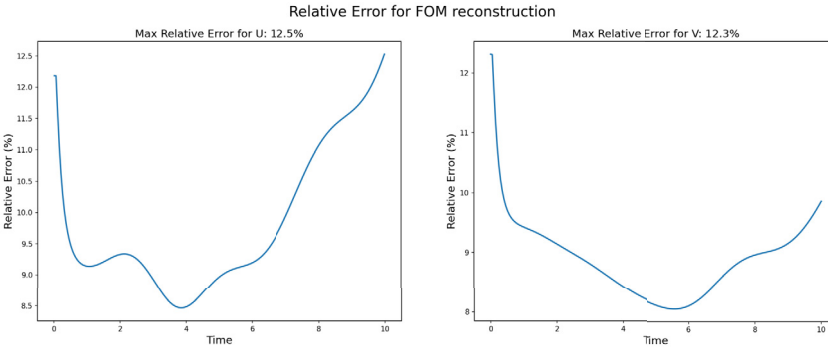


FIGURE 10 Latent space trajectories of the reaction-diffusion system with 10% noise using POD.



(a) Relative error of the system simulation with 10% noise



(b) Relative error of the system simulation with 100% noise

FIGURE 11 Relative reconstruction error in simulations.

dimension of 15. This process yields the latent space trajectories for both u and v, as illustrated in Fig. 10.

We calculate the relative error of the simulation results for both u and v. Notably, with a 10% noise level, we attain a remarkable relative error below

3%. Furthermore, when subject to 100% noise, the relative error remains below 13% as illustrated in Fig. 11. Particularly, with WLaSDI, the simulation is completed about 250 times faster compared to numerically solving the equation. Additional details regarding WLaSDI, along with a comprehensive comparison between the weak and strong form when applying to projection-based reduced order modeling, can be found in Tran et al. (2024).

6 Conclusions

Weak form methods are a novel class of scientific machine learning algorithms. They have found success in a wide range of areas including sparse regression-based automated discovery of governing equations, direct parameter estimation, and reduced order models. In all cases the use of the weak form allows improves robustness of the method to noisy state measurements.

Acknowledgments

This work was supported in part by NSF grant MCB-2054085, DOE grant DE-SC0023346, and NIH grant R35GM149335.

References

Bortz, David M., Messenger, Daniel A., Dukic, Vanja, 2023. Direct estimation of parameters in ODE models using WENDy: weak-form estimation of nonlinear dynamics. Bull. Math. Biol. 85 (110).

Brunton, Steven L., Proctor, Joshua L., Kutz, J. Nathan, 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. 113 (15), 3932–3937.

Chen, Hao, 2023. Data-driven sparse identification of nonlinear dynamical systems using linear multistep methods. Calcolo 60 (1), 11.

Chen, Xiaoli, Yang, Liu, Duan, Jinqiao, Karniadakis, George Em, 2021. Solving inverse stochastic problems from discrete particle observations using the Fokker–Planck equation and physics-informed neural networks. SIAM J. Sci. Comput. 43 (3), B811–B830.

Diaconis, Persi, 1988. Bayesian numerical analysis. In: Gupta, S.S., Berger, James O. (Eds.), Statistical Decision Theory and Related Topics IV, vol. 1, pp. 163–175.

Fasel, U., Kutz, J.N., Brunton, B.W., Brunton, S.L., 2022. Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control. Proc. R. Soc., Math. Phys. Eng. Sci. 478 (2260), 20210904.

Gurevich, Daniel R., Reinbold, Patrick A.K., Grigoriev, Roman O., 2019. Robust and optimal sparse regression for nonlinear pde models. Chaos 29 (10), 103113.

Houston, Paul, Schwab, Christoph, Süli, Endre, 2002. Discontinuous hp-finite element methods for advection-diffusion-reaction problems. SIAM J. Numer. Anal. 39 (6), 2133–2163.

Jorgensen, Murray, 2012. Iteratively reweighted least squares. In: El-Shaarawi, Abdel H., Piegorsch, Walter W. (Eds.), Encyclopedia of Environmetrics, first edition. Wiley.

Lagergren, John H., Nardini, John T., Lavigne, G. Michael, Rutter, Erica M., Flores, Kevin B., 2020. Learning partial differential equations for biological transport models from noisy spatio-temporal data. Proc. R. Soc. A 476 (2234), 20190800.

Lax, P.D., Milgram, A.N., 1955. IX. Parabolic Equations. Annals of Mathematical Studies, vol. 33. Princeton University Press, pp. 167–190.

Liu, Wing Kam, Li, Shaofan, Park, Harold S., 2022. Eighty years of the finite element method: birth, evolution, and future. Arch. Comput. Methods Eng. 29 (6), 4431–4453.

Messenger, Daniel A., Bortz, David M., 2021a. Weak SINDy for partial differential equations. J. Comput. Phys. 443, 110525.

Messenger, Daniel A., Bortz, David M., 2021b. Weak SINDy: Galerkin-based data-driven model selection. Multiscale Model. Simul. 19 (3), 1474–1497.

Messenger, Daniel A., Bortz, David M., 2022a. Asymptotic consistency of the wsindy algorithm in the limit of continuum data. arXiv preprint. arXiv:2211.16000.

Messenger, Daniel A., Bortz, David M., 2022b. Learning mean-field equations from particle data using wsindy. Phys. D: Nonlinear Phenom. 439, 1474–1497, 133406.

Messenger, Daniel A., Burby, Joshua W., Bortz, David M., 2023. Coarse-Graining Hamiltonian Systems Using WSINDy. arXiv preprint. arXiv:2310.05879.

Messenger, Daniel A., Dall'Anese, Emiliano, Bortz, David M., 2022a. Online weak-form sparse identification of partial differential equations. In: Proceedings of The Third Mathematical and Scientific Machine Learning Conference, vol. 190, pp. 241–256.

Messenger, Daniel A., Wheeler, Graycen E., Liu, Xuedong, Bortz, David M., 2022b. Learning anisotropic interaction rules from individual trajectories in a heterogeneous cellular population. J. R. Soc. Interface 19 (195), 20220412.

Pantazis, Yannis, Tsamardinos, Ioannis, 2019. A unified approach for sparse dynamical system inference from temporal measurements. Bioinformatics 35 (18), 3387–3396.

Rudy, Samuel H., Brunton, Steven L., Proctor, Joshua L., Kutz, J. Nathan, 2017. Data-driven discovery of partial differential equations. Sci. Adv. 3 (4), e1602614.

Schaeffer, Hayden, McCalla, Scott G., 2017. Sparse model selection via integral terms. Phys. Rev. E 96 (2), 023302.

Schwartz, Laurent, 1950. Théorie Des Distributions, volume I. Hermann et Cie, Paris, France.

Shinbrot, Marvin, 1954. On the analysis of linear and nonlinear dynamical systems for transient-response data. Technical Report NACA TN 3288. Ames Aeronautical Laboratory, Moffett Field, CA.

Tran, April, He, Xiaolong, Messenger, Daniel A., Choi, Youngsoo, Bortz, David M., 2024. Weak-form latent space dynamics identification. Comput. Methods Appl. Mech. Eng. 356, 116998.

Vershynin, Roman, 2018. High-Dimensional Probability: An Introduction with Applications in Data Science, vol. 47. Cambridge University Press.

Wang, Zhenlin, Huan, Xun, Garikipati, Krishna, 2019. Variational system identification of the partial differential equations governing the physics of pattern-formation: inference under varying fidelity and noise. Comput. Methods Appl. Mech. Eng. 356, 44–74.