

Fourier Weak SINDy: Spectral Test Function Selection for Robust Model Identification

Zhiheng Chen

ZC548@CORNELL.EDU

Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York, USA, 14850

Urban Fasel

U.FASEL@IMPERIAL.AC.UK

Department of Aeronautics, Imperial College London, London, United Kingdom

Anastasia Bizyaeva

ANASTASIAB@CORNELL.EDU

Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York, USA, 14850

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

We introduce Fourier Weak SINDy, a minimal noise-robust and interpretable derivative-free equation learning method that combines weak-form sparse equation learning with spectral density estimation for data-driven test function selection. By using orthogonal sinusoidal test functions inspired by their prevalence in Modulating Function-based system identification, the weak-form sparse regression problem reduces to a regression over Fourier coefficients. Dominant frequencies are then selected via multitaper estimation of the frequency spectrum of the data. This formulation unifies weak-form learning and spectral estimation within a compact and flexible framework. We illustrate the effectiveness of this approach in numerical experiments across multiple chaotic and hyperchaotic ODE benchmarks.

Keywords: equation learning, nonlinear system identification, sparse regression, nonlinear dynamical systems, spectral density estimation

1. Introduction

In this paper, we present a minimal interpretable framework for noise-robust derivative-free equation learning that combines recent advances in sparse system identification with classical signal processing tools to address the challenge of data-driven test function selection in the algorithm design. In recent years, sparse regression based methods for data-driven modeling have received significant attention. Among these, the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm is a popular data-driven framework for learning Ordinary Differential Equations (ODEs), that uses sparsity-promoting optimization to identify unknown model coefficients while avoiding overfitting (Brunton et al., 2016). SINDy has been applied to equation learning problems in different domains from fluid dynamics to disease modeling (Fukami et al., 2021; Delabays et al., 2025; Horrocks and Bauch, 2020), and has found many extensions including learning partial differential equations (Rudy et al., 2017), simultaneously learning coordinates and differential equations (Champion et al., 2019; Bakarji et al., 2023), and quantifying uncertainty in the learned model structure and coefficients Zhang and Lin (2018); North et al. (2022); Niven et al. (2024); Fung et al. (2025).

Measurement noise and limited data are key challenges for learning models from data, motivating extensions of the SINDy framework such as ensembling (Fasel et al., 2022), denoising signals (Kaheman et al., 2022), and weak-form reformulations (Schaeffer and McCalla, 2017; Reinbold et al., 2020; Messenger and Bortz, 2021; Messenger et al., 2024; Stephany and Earls, 2024). The

weak-form or integral-form family of methods are particularly promising noise-robust variants of SINDy, as they bypass the need for estimation of derivative information from noise corrupted data, which makes the original SINDy framework fragile. Weak-form methods instead rely on integrating data against user-defined analytic smooth test functions and their derivatives. Success of these methods therefore relies on selecting and tuning a good test function basis given characteristics of the collected data, to which significant attention has been devoted in recent work. For example, [Messenger and Bortz \(2021\)](#) propose adaptive location and radius selections for compactly-supported “bump”-shaped test functions based on approximate error covariance, the width-at-half-max parameter, and dominant wave modes; [Stephany and Earls \(2024\)](#) propose an optimization framework for test function tuning for weak-form PDE learning; [Bortz et al. \(2023\)](#) start with compactly supported test functions and orthogonalize them using SVD, with the orthogonalized test functions resembling local Fourier basis or wavelets; [Gurevich et al. \(2019\)](#) provide a detailed heuristic discussion of test function selection based on the frequency content of the data.

In modern weak-form sparse equation learning formulations, “bump”-shaped test functions are a popular design choice due to their compact support and interpretable parameters. However, this choice introduces significant degrees of freedom into the test function selection problem, as the support region boundaries as well as the shape parameters of the test functions become optimization parameters. Furthermore, such test functions are not orthogonal, which means redundant information in the signal is reused in different components of the regression problem. In the work presented in this paper, we will illustrate that a simple test function basis of orthogonal sinusoids is an interpretable and robust choice for weak-form learning. In this formulation, we take inspiration from classical Modulating Function Method system identification literature ([Shinbrot, 1957](#)), in which orthogonal test functions (or modulating functions) have been widely used for system identification ([Pearson and Lee, 1985](#); [Pearson and Pan, 1990](#); [Pearson, 1995](#)). The utility of orthogonal test functions has also recently been noted in the weak SINDy literature with data-driven constructions ([Bortz et al., 2023](#)) and derivation of general error bounds ([Russo and Laiu, 2024](#)). Furthermore, we will show that with sinusoidal test functions, classical spectral density estimation (SDE) presents a natural strategy for data-driven selections of test function frequencies. Common approaches to SDE include nonparametric methods such as Bartlett method and Welch method ([Stoica et al., 2005](#)); a more advanced approach with higher resolutions and lower variances is the multitaper method ([Thomson, 2005](#)). Due to its strong frequency specificity, the multitaper method in particular is popular across disciplines for the spectral analysis of complex time series data, from seismic data analysis to brain wave analysis ([Park et al., 1987](#); [van Vugt et al., 2007](#); [Babadi and Brown, 2014](#)). We will utilize multitaper to inform the selection of a sinusoidal test function basis.

In this paper, we present the following contributions. First, we introduce Fourier Weak SINDy, a minimal method that combines weak-form sparse regression with spectral density estimation to learn parsimonious sparse ODE models from noisy measurement data. A high-level graphical summary of the method can be found in [Figure 1](#). We show that under the choice of sinusoidal test functions, the Weak SINDy sparse regression problem simplifies to a regression over Fourier series coefficients, which are efficiently computable with the Fast Fourier Transform (FFT). The choice of coefficients to regress over is made in a data-driven way by selecting dominant frequency modes in the data using multitaper estimation. Second, we illustrate effectiveness of this minimal approach in simulation studies. In particular, we provide numerical evidence that our method outperforms baseline SINDy and Weak SINDy methods across several chaotic and hyperchaotic ODE benchmarks for varying levels of measurement noise. Importantly, the SINDy and Weak SINDy baselines we

compare against follow the implementations in the standard PySINDy package (Kaptanoglu et al., 2022). We leave principled comparisons against Weak SINDy variants with adaptive hyperparameter tuning, for example as in Messenger and Bortz (2021), to future work. Finally, we investigate the sensitivity of the method to the size of test function dictionary and the bandwidth of the spectral density estimation, illustrating that a reasonably small test function dictionary and bandwidth window yields robust performance on the low-dimensional ODE benchmarks. Our simulation code is available in a shared [Github repository](#).

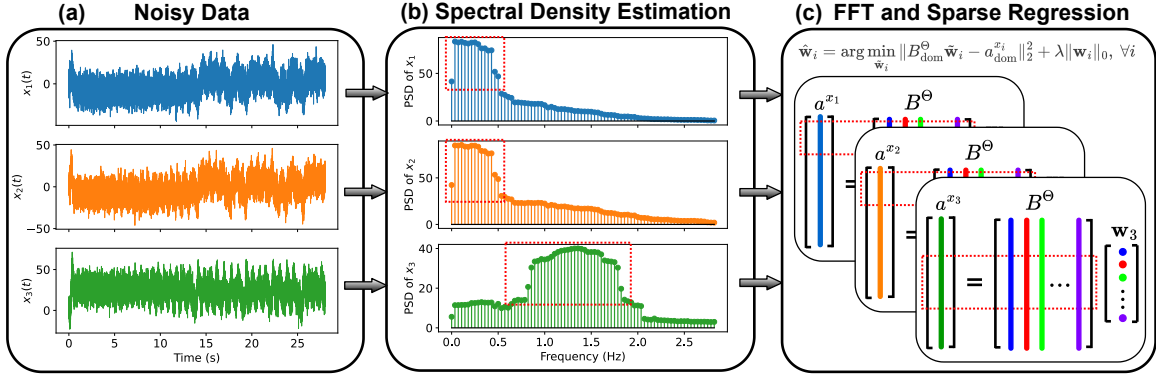


Figure 1: Graphical summary of proposed Fourier Weak SINDy method for sparse equation learning described in detail in Section 3. First, multitaper spectral density estimation is used to identify dominant temporal frequencies in the observed noisy signals. Second, spectral sparse regression problem is solved to learn model coefficients in a selected function dictionary, leveraging the identified dominant Fourier frequencies as test functions in a weak-form system identification problem.

This paper is structured as follows. Section 2 reviews the fundamental formulations and limitations of current SINDy and weak SINDy frameworks; Section 3 presents our Fourier weak SINDy method, with mathematical derivations for the frequency-domain representation and the frequency selections based on spectral density estimations; Section 4 shows numerical experiment results, where Fourier weak SINDy is compared with standard SINDy and weak SINDy on equation learning tasks of four benchmark systems; Section 5 discusses limitations and future work.

2. Background

Consider the problem of learning the structure of the governing equations for a finite-dimensional autonomous dynamical system given noisy observations of its state,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\xi}(t) \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\dot{\mathbf{x}}$ is the time derivative, $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, and $\boldsymbol{\xi}(t) \in \mathbb{R}^n$ is a vector of additive measurement noise with components $\xi_i(t) \sim \mathcal{N}(0, \sigma^2)$ independent across state dimensions and time. We assume the output is sampled at regular time intervals $t \in \{t_1, \dots, t_k\}$, $t_{i+1} - t_i = \Delta t$, with the collected data snapshots summarized in the matrix

$$Y = [\mathbf{y}(t_1) \quad \dots \quad \mathbf{y}(t_k)]^T \in \mathbb{R}^{k \times n}. \quad (2)$$

2.1. Sparse Identification of Nonlinear Dynamics (SINDy)

The Sparse Identification of Nonlinear Dynamics (SINDy) method for equation learning has received significant interdisciplinary attention (Brunton et al., 2016, 2025). SINDy combines classical least-squares system identification (Åström and Eykhoff, 1971; Unbehauen and Rao, 1990) with a sparsity-promoting sequentially thresholded least squares solver to identify the coefficients of nonlinear terms in the function $\mathbf{f}(\mathbf{x})$ in (1) from a large dictionary of candidate functions, under the assumption that $\mathbf{f}(\mathbf{x})$ is linear-in-coefficients and sparse in the selected dictionary.

To set up the sparse identification problem, define m dictionary functions $\{\theta_1, \dots, \theta_m\}$ where each $\theta_i : \mathbb{R}^n \rightarrow \mathbb{R}$. Common dictionary choices include polynomials up to degree d , i.e. $\{\theta_j(\mathbf{x})\}_{j=1}^m = \{1, x_1, \dots, x_n, x_1^2, x_1x_2, \dots, x_n^d\}$, or other function bases such as trigonometric functions. It is assumed that each component $f_i(\mathbf{x})$ of the vector field $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_n(\mathbf{x})]^T$ admits a representation

$$f_i(\mathbf{x}(t)) = \sum_{j=1}^m w_{ij} \theta_j(\mathbf{x}(t)) = \Theta(\mathbf{x}(t)) \mathbf{w}_i \quad (3)$$

where $\mathbf{w}_i \in \mathbb{R}^m$ is a coefficient vector and $\Theta(\mathbf{x}(t)) = [\theta_1(\mathbf{x}(t)) \ \dots \ \theta_m(\mathbf{x}(t))] \in \mathbb{R}^{1 \times m}$. The true coefficient matrix $W = [\mathbf{w}_1 \ \dots \ \mathbf{w}_n] \in \mathbb{R}^{m \times n}$ parametrizes the right hand side of (1) with respect to the function dictionary as $\mathbf{f}(\mathbf{x}(t)) = \Theta(\mathbf{x}(t))W$. The objective is to recover the unknown coefficients W from the observed snapshots (2) under the assumption that each \mathbf{w}_i is sparse, i.e. each function component $f_i(\mathbf{x}(t))$ comprises only a small number of dictionary terms.

Given time series Y of state measurements defined in (2), an estimate of the corresponding time derivatives of the state $\dot{X} = [\dot{\mathbf{x}}(t_1) \ \dots \ \dot{\mathbf{x}}(t_k)]^T \in \mathbb{R}^{k \times n}$ is constructed, for example using finite-difference approximations or polynomial smoothing. By stacking evaluations of the dictionary functions at each snapshot, we can construct the dictionary matrix

$$\Theta(Y) = \begin{bmatrix} \Theta(\mathbf{y}(t_1)) \\ \vdots \\ \Theta(\mathbf{y}(t_k)) \end{bmatrix} = \begin{bmatrix} \theta_1(\mathbf{y}(t_1)) & \dots & \theta_m(\mathbf{y}(t_1)) \\ \vdots & \ddots & \vdots \\ \theta_1(\mathbf{y}(t_k)) & \dots & \theta_m(\mathbf{y}(t_k)) \end{bmatrix} \in \mathbb{R}^{k \times m}. \quad (4)$$

Approximation of the unknown coefficient matrix W can be formulated as the least squares problem

$$\widehat{W} = \arg \min_{\tilde{W}} \left\| \dot{X} - \Theta(Y)\tilde{W} \right\|_2^2. \quad (5)$$

The regression problem (5) is then solved using a sparsity promoting method such as sequentially thresholded least squares or sequentially thresholded ridge regression as proposed in (Brunton et al., 2016; Rudy et al., 2017); that is, we iteratively solve (5) using standard least squares or ridge regression, and at each iteration setting to zero coefficients whose magnitudes fall below a specified threshold $\eta > 0$. After K iterations, the resulting coefficient matrix \widehat{W} identifies active dictionary terms in each component of the vector field (3).

2.2. Weak SINDy and Test Function Selection

While SINDy is an efficient and interpretable framework for model discovery, its reliance on derivative estimates \dot{X} from noisy measurements limits its robustness. In practice, SINDy coefficient estimates become highly inaccurate even with reasonably large signal-to-noise ratios in the data matrix

(2). To mitigate this noise sensitivity, a number of recent research efforts proposed integral-form or weak-form reformulations of SINDy, which bypass the need for estimating derivatives (Schaeffer and McCalla, 2017; Gurevich et al., 2019; Reinbold et al., 2020; Messenger and Bortz, 2021; Bortz et al., 2023; Stephany and Earls, 2024). In the Weak SINDy formulation, each component $f_i(\mathbf{x})$ of (1) is projected onto a compactly-supported test function $\phi_j \in C_c^1([t_a, t_b])$ and integrated over its support region, $\int_{t_a}^{t_b} \dot{x}_i(t)\phi_j(t)dt = \int_{t_a}^{t_b} f_i(\mathbf{x}(t))\phi_j(t)dt$. Imposing that test functions vanish at the end points of the time interval, i.e. $\phi_j(t_a) = \phi_j(t_b) = 0$, and applying integration-by-parts yields

$$-\int_{t_a}^{t_b} x_i(t)\phi_j'(t)dt = \int_{t_a}^{t_b} f_i(\mathbf{x}(t))\phi_j(t)dt. \quad (6)$$

In discrete time, the integrals in (6) are approximated using numerical quadrature. The matrices of evaluations of $\phi_j(t)$ and its derivative $\phi_j'(t)$ at the sampled time points are defined as $\Phi_j = [\phi_j(t_1) \dots \phi_j(t_k)] \in \mathbb{R}^{1 \times k}$ and $\Phi_j' = [\phi_j'(t_1) \dots \phi_j'(t_k)] \in \mathbb{R}^{1 \times k}$. Defining $X \in \mathbb{R}^{k \times n}$ as the data snapshot matrix and leveraging the library matrix expansion (3) along with the dictionary matrix definition (4), the integral expression (6) becomes $-\Delta t \Phi_j' X = \Delta t \Phi_j \Theta(X) W$. Collecting evaluations of p distinct test functions into the matrices $\Phi \in \mathbb{R}^{p \times k}$ and $\Phi' \in \mathbb{R}^{p \times k}$, and incorporating noisy data observations (2), coefficients W of the ODE are then estimated by solving

$$\widehat{W} = \arg \min_{\tilde{W}} \left\| \Delta t \Phi \Theta(Y) \tilde{W} + \Delta t \Phi' Y \right\|_2^2. \quad (7)$$

As in the standard SINDy formulation, coefficient sparsity is enforced by solving (7) using least squares or ridge regression with sequential thresholding.

A key design choice that significantly affects the performance of weak-form methods is the form of test function $\phi_j(t)$. Grounded in the definition of a weak derivative, test functions ϕ_j are often constrained to be compactly supported over the observation interval. Typical test functions have the form of a localized ‘‘bump’’, for example Reinbold et al. (2020) utilize functions of the form

$$\phi_j(t) = (1 - \underline{t}^2)^q \quad (8)$$

where $q \in \mathbb{Z}^+$ is a hyperparameter (larger q gives narrower bumps around the subdomain center), and \underline{t} is the dimensionless time obtained via $\underline{t} = (t - t_j)/H_t$ with t_j being the center of the j -th subdomain and H_t being the duration of each subdomain. Optimizing the performance of weak SINDy then requires strategic selections of the sharpness and support interval boundaries of the test functions. For example, Messenger and Bortz (2021); Bortz et al. (2023) present an adaptive-grid strategy that places more test functions at locations where large changes in the dynamics occur, where the gradient is approximated by formulating and solving a nonlinear root finding problem; then they select hyperparameters and support intervals based on the approximate error covariance, the width-at-half-max parameter, and optionally the dominant wave modes.

3. Fourier Weak SINDy

The utility of orthogonal test functions has been explored in several prior works. For example, Bortz et al. (2023) construct test functions by orthogonalizing smooth bump functions via Singular Value Decomposition (SVD). Moreover, in classical systems and control literature, such as Pearson and Lee (1985); Pearson and Pan (1990); Pearson (1995), Fourier-mode test functions (also known as

modulating functions) are used for identification of linear systems, without requiring them to be compactly supported, and only requiring that they should vanish at domain boundaries. Motivated by these precedents, in this section we introduce sinusoidal test functions into the weak SINDy framework and illustrate that the resulting learning problem reduces to a sparse regression over Fourier coefficients. This reformulation also reinterprets test function selection as the selection of informative Fourier frequencies, which can be addressed using classical signal processing methods for spectral density estimation. Together, this formulation yields a noise-robust and interpretable spectral approach to sparse equation learning.

3.1. Sinusoidal Test Functions

Suppose the data we collect has duration T (i.e., $T = t_k - t_1$), then define the ℓ -th sinusoidal test function and its derivative, where $\ell \in \mathbb{Z}^+$, as

$$\phi_\ell(t) = \sin\left(\frac{2\pi\ell}{T}t\right), \quad \phi'_\ell(t) = \frac{2\pi\ell}{T} \cos\left(\frac{2\pi\ell}{T}t\right). \quad (9)$$

Since $\phi_\ell(0) = \phi_\ell(T) = 0$ for all $\ell \in \mathbb{Z}^+$, the weak form equation (6) becomes

$$-\frac{2\pi\ell}{T} \int_0^T x_i(t) \cos\left(\frac{2\pi\ell}{T}t\right) dt = \int_0^T f_i(\mathbf{x}(t)) \sin\left(\frac{2\pi\ell}{T}t\right) dt. \quad (10)$$

Let $x_i(t)$ and $f_i(t)$ defined over $[0, T]$ be expanded as Fourier series

$$x_i(t) = a_0^{x_i} + \sum_{\ell=1}^{\infty} \left(a_\ell^{x_i} \cos\left(\frac{2\pi\ell}{T}t\right) + b_\ell^{x_i} \sin\left(\frac{2\pi\ell}{T}t\right) \right), \quad (11a)$$

$$f_i(x(t)) = a_0^{f_i} + \sum_{\ell=1}^{\infty} \left(a_\ell^{f_i} \cos\left(\frac{2\pi\ell}{T}t\right) + b_\ell^{f_i} \sin\left(\frac{2\pi\ell}{T}t\right) \right). \quad (11b)$$

Since the Fourier basis are mutually orthogonal, substituting (11) into (10) lets most of the inner products vanish, giving

$$-\frac{2\pi\ell}{T} a_\ell^{x_i} = b_\ell^{f_i}, \quad \text{for all } \ell \in \mathbb{Z}^+. \quad (12)$$

Note that the $a_\ell^{x_i}$ and $b_\ell^{f_i}$ coefficients can be efficiently computed using FFT of the data matrix Y and the dictionary matrix $\Theta(Y)$.

3.2. Dictionary Function Expansion

In the equation learning problem, we do not know the exact form of f_i and instead expand it in a dictionary of functions $\Theta(\mathbf{x}) = [\theta_1(\mathbf{x}) \ \dots \ \theta_m(\mathbf{x})]$, with the relation described in (3). Let the Fourier series expansion of $\theta_j(\mathbf{x}(t))$ defined over $[0, T]$ be

$$\theta_j(\mathbf{x}(t)) = a_0^{\theta_j} + \sum_{\ell=1}^{\infty} \left(a_\ell^{\theta_j} \cos\left(\frac{2\pi\ell}{T}t\right) + b_\ell^{\theta_j} \sin\left(\frac{2\pi\ell}{T}t\right) \right). \quad (13)$$

By linearity of the Fourier operator, the right-hand side of (12) can then be written as

$$b_\ell^{f_i} = \sum_{j=1}^m w_{ij} b_\ell^{\theta_j}. \quad (14)$$

Define $B_\ell^\ominus = [b_\ell^{\theta_1} \ \dots \ b_\ell^{\theta_m}]$; then $b_\ell^{f_i} = B_\ell^\ominus \mathbf{w}_i$, and (12) becomes

$$-\frac{2\pi\ell}{T} a_\ell^{x_i} = B_\ell^\ominus \mathbf{w}_i. \quad (15)$$

Since for each j , $b_\ell^{\theta_j}$ for all ℓ up to the index for Nyquist frequency can be efficiently computed using one FFT, this formulation can be significantly faster to implement than quadrature-based approaches, where an evaluation of quadrature is needed for each test function.

3.3. Frequency Selection Based on Spectral Density Estimation

The multitaper method provides a low-variance, low-leakage, and high-resolution approach to spectral density estimation (SDE) compared to simple single-taper approaches Thomson (2005). Therefore, we use the multitaper method to perform SDE for trajectory data in order to select informative spectral test function frequencies. A brief overview of the multitaper method with Slepian sequences is available in the supplement.

For each f_i , we apply the multitaper method to estimate the power spectral density of $x_i(t)$ from noisy measurements, and select frequencies at which the estimated power is highest, i.e. the dominant frequencies. At dominant frequencies, the signal-to-noise ratio is highest, making the corresponding spectral content the most reliable basis for system identification from noisy data. Suppose we pick K dominant frequencies in total, and define

$$a_{\text{dom}}^{x_i} = \left[-\frac{2\pi}{T} a_1^{x_i} \ \dots \ -\frac{2\pi K}{T} a_K^{x_i} \right]^T, \quad B_{\text{dom}}^\ominus = \left[(B_1^\ominus)^T \ \dots \ (B_K^\ominus)^T \right]^T. \quad (16)$$

Then we can obtain, and solve using sequentially thresholded methods, the least squares problem

$$\hat{\mathbf{w}}_i = \arg \min_{\tilde{\mathbf{w}}_i} \| B_{\text{dom}}^\ominus \tilde{\mathbf{w}}_i - a_{\text{dom}}^{x_i} \|_2^2. \quad (17)$$

This approach to test function selection is interpretable in a physically meaningful sense, since the selected frequencies correspond directly to the dominant modes of the observed signal. The choice of test functions is directly grounded in the spectral structure of the data, rather than indirectly through abstract optimization criteria, and utilizes classical signal processing methods.

4. Simulation Results

We implement the Fourier Weak SINDy framework in Python, testing it on simulated trajectories of several benchmark nonlinear dynamical systems with varying measurement noise levels. The trajectory data have durations of 10 s and are evenly sampled with 1000 samples per second. We follow the noise level definition and error metrics in Messenger and Bortz (2021). Specifically, the noise ratio σ_{NR} is defined using

$$\sigma = \sigma_{NR} \frac{\|X\|_F}{\sqrt{kn}} \quad (18)$$

where $\|X\|_F$ is the Frobenius norm of the trajectory data matrix $X \in \mathbb{R}^{k \times n}$, and the noise level is then $\sigma_{NR} \times 100\%$. We use two metrics to evaluate the error of equation learning results; the first metric is the relative coefficient error $E_2(\hat{W})$, defined as

$$E_2(\hat{W}) = \frac{\|\hat{W} - W\|_F}{\|W\|_F}. \quad (19)$$

The second metric is the true positive ratio (TPR)

$$\text{TPR}(\hat{W}) = \frac{\text{TP}(\hat{W})}{\text{TP}(\hat{W}) + \text{FP}(\hat{W}) + \text{FN}(\hat{W})} \quad (20)$$

where $\text{TP}(\hat{W})$ is the number of correctly identified nonzero terms, $\text{FP}(\hat{W})$ is the number of falsely identified nonzero terms, and $\text{FN}(\hat{W})$ is the number of terms that are falsely identified as having a coefficient of zero.

In the equation learning of all benchmark systems using Fourier weak SINDy, we use the sequentially thresholded ridge regression with a sparsity threshold of 0.5 and a ridge regularization value of 0.001. We implement two frequency selection methods for Fourier weak SINDy – the first method is to use the first 500 frequencies of the FFT result (i.e., using a sweep of frequencies up to 50 Hz, $\frac{1}{10}$ of the Nyquist frequency), and the second method is to use the top 100 dominant frequencies from the SDE. We compare the results of Fourier weak SINDy with those of standard SINDy and weak SINDy methods in the PySINDy package [Kaptanoglu et al. \(2022\)](#), where we choose sparse relaxed regularized regression to obtain sparse coefficient matrices, and use 1000 subdomains for the weak SINDy in PySINDy (i.e., 1000 test functions of the form (8)).

We perform our first numerical experiment on the Lorenz system with library terms up to second-order polynomials. We set the initial conditions of the simulation to $\mathbf{x}_0 = [20 \ 12 \ -30]^T$. As shown in Fig. 2, we first compare the performances of SINDy (PySINDy), weak SINDy (PySINDy), and Fourier weak SINDy based on both the frequency sweep and the spectral density estimation methods, with noise ratios ranging from 0.0001% to 100% and 20 noise instances per noise level. Moreover, we investigate the effect of choosing different numbers of dominant frequencies for the Fourier weak SINDy at different noise levels (the dominant frequencies are extracted using the FFT on the clean data, to eliminate the effect of SDE error and focus on the effect of frequency numbers), as well as how the bandwidth of the multitaper SDE influences equation learning results.

As shown in Fig. 2(a) for the Lorenz system benchmark, at 100% noise level, SINDy had a median coefficient error at the order of 10 and a median TPR near 0.2. This illustrates that the classical SINDy approach is not providing meaningful equation learning results at high measurement noise. In contrast, the Fourier weak SINDy approach with SDE achieved a median coefficient error on the order of 0.1 and a median TPR near 0.7. This implies that even with significant measurement noise, Fourier weak SINDy can still reveal the core structure of the model with reasonably small coefficient error. Furthermore, up through a 25% noise level, Fourier Weak SINDy had a TPR of 1 across all of the numerical experiments, meaning that the sparse regression problem yielded no false positives or false negatives. This is significant improvement over the SINDy and Weak SINDy baselines, and over the Fourier Weak SINDy method with a sweep frequency selection. In addition to coefficient error and TPR, we evaluated simulated trajectory error for the Lorenz system, illustrated in a figure available in the Supplement. We find that Fourier Weak SINDy outperforms the baseline methods, and is less likely to learn unstable dynamics.

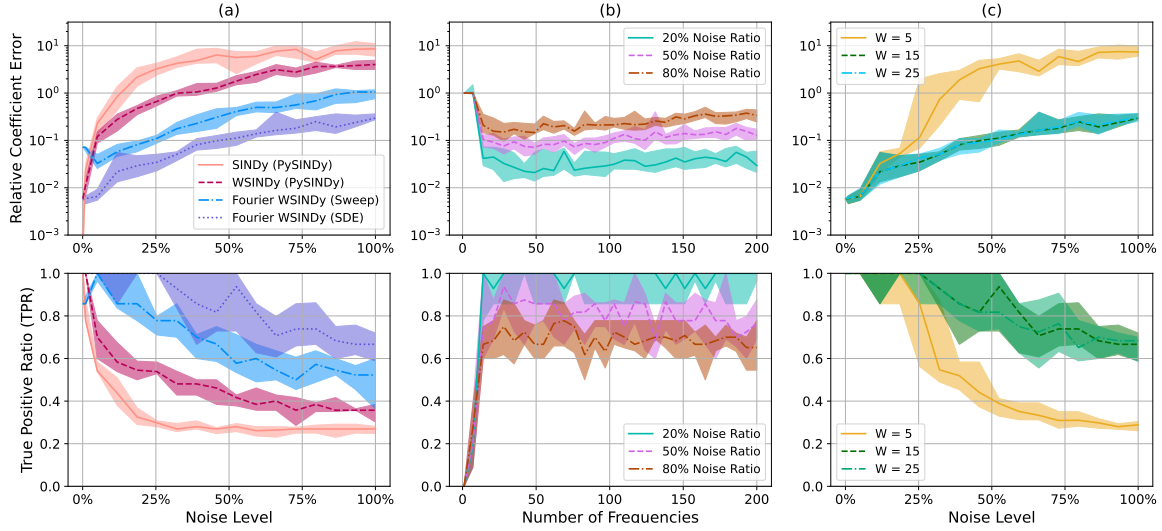


Figure 2: Lorenz system equation learning results with second-order polynomial library. Median and quartiles of the relative coefficient error and TPR of (a) four different SINDy and weak SINDy methods at varying noise levels, (b) Fourier weak SINDy with SDE and different numbers of dominant frequencies chosen, and (c) Fourier weak SINDy with different bandwidth settings (in Hz) for the multitaper SDE.

Fig. 2(b) illustrates the performance Fourier weak SINDy against the number of test functions being selected. Once the number of test functions is more than 30, the coefficient error and true positive ratio does not vary significantly as the number of test functions increases, though there is a slight decline in performance as test functions of higher frequencies are included. The slight decline in performance is likely caused by the fact that the frequency content of the signal is mostly in the top 50 to 100 dominant frequencies, and thus the components projected onto the less-dominant test functions contain more information from measurement noise that is not dynamically meaningful. Fig. 2(c) reveals the performance of Fourier weak SINDy against the selection of the bandwidth of the multitaper SDE. The performance improves with the size of the bandwidth at first, but the benefit saturates as the bandwidth increases.

We also perform numerical experiments on the Lotka-Volterra equations, the hyperchaotic Lorenz system, and the hyperchaotic Jha system, all with second-order polynomial libraries; the equation learning results are shown in Fig. 3. The governing equations of all four benchmark systems, parameters and initial conditions used for benchmarking, and equation learning results with higher-order polynomial libraries are available in the supplement. As shown in Fig. 3, Fourier weak SINDy with SDE typically improves on, or at worst performs comparably to, the baseline methods on these systems according to the equation error and TPR metrics.

5. Discussion

In this paper, we showed that combining weak-form equation learning over orthogonal sinusoidal test functions with classical signal processing techniques for the selection of test function frequen-

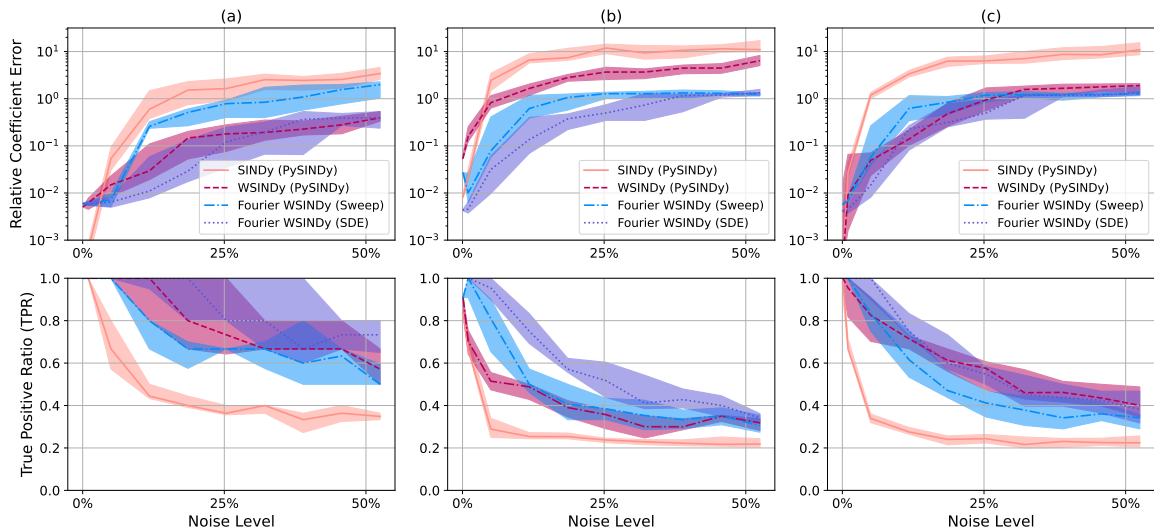


Figure 3: Equation learning results of four different SINDy and weak SINDy methods for (a) Lotka-Volterra equations, (b) hyperchaotic Lorenz system, and (c) hyperchaotic Jha system.

cies to match maximally energetic modes in the data, presents an interpretable, robust, and effective approach to equation learning. We presented numerical evidence that Fourier weak SINDy shows either improved or comparable performance in equation learning tasks across four chaotic and hyperchaotic benchmark systems, compared to baseline methods including the “bump” weak SINDy method in the PySINDy package. Moreover, the additional hyperparameters introduced by the method over the baselines, i.e. the number of dominant frequencies to be included in the regression problem and the bandwidth selection for the multitaper SDE, are interpretable and were shown numerically to have a saturating effect on method performance, providing practitioners with a simple and intuitive strategy for tuning.

There are many promising directions of future work. For example, the current FFT implementations of Fourier weak SINDy are limited to uniformly sampled trajectory data. Therefore, a future direction is to experiment with Nonuniform Fast Fourier Transform to perform equation learning using data with irregular time samples. Moreover, an assumption made in this paper is that the measurement noise is additive white Gaussian noise. Although this assumption holds for many equation learning tasks, noise profiles beyond white noise are common in practice. In these cases, Fourier weak SINDy can be further augmented by more advanced spectral density estimation methods, such as the adaptive multitaper method and the Capon estimator. Additionally, existing techniques for improving the accuracy and noise-robustness of SINDy, such as ensembling and Bayesian coefficient selection strategies, can be incorporated into the current framework to further improve accuracy and robustness. Finally, the current framework utilizes spectral density estimation for test function selection; while the multitaper method effectively captures the dominant frequencies in terms of energy, dynamically meaningful information may be contained in low-amplitude modes of the signal. In this case, alternative test function selection strategies may be more effective, for example selection based on the Fourier Phase Index metric proposed by [Aguilar-Hernández et al. \(2024\)](#).

Acknowledgments

The authors are grateful to Daniel Messenger for helpful literature pointers to recent work on weak SINDy with orthogonal test functions.

References

- Alberto Isaac Aguilar-Hernández, David Michel Serrano-Solis, Wady A Ríos-Herrera, José Fernando Zapata-Berruecos, Gloria Vilaclara, Gustavo Martínez-Mekler, and Markus F Müller. Fourier phase index for extracting signatures of determinism and nonlinear features in time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(1), 2024.
- Karl Johan Åström and Peter Eykhoff. System identification—a survey. *Automatica*, 7(2):123–162, 1971.
- Behtash Babadi and Emery N Brown. A review of multitaper spectral analysis. *IEEE Transactions on Biomedical Engineering*, 61(5):1555–1564, 2014.
- Joseph Bakarji, Kathleen Champion, J Nathan Kutz, and Steven L Brunton. Discovering governing equations from partial measurements with deep delay autoencoders. *Proceedings of the Royal Society A*, 479(2276):20230422, 2023.
- David M Bortz, Daniel A Messenger, and Vanja Dukic. Direct estimation of parameters in ode models using wendy: Weak-form estimation of nonlinear dynamics. *Bulletin of Mathematical Biology*, 85(11):110, 2023.
- Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.
- Steven L Brunton, Nicholas Zolman, J Nathan Kutz, and Urban Fasel. Machine learning for sparse nonlinear modeling and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 8, 2025.
- Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.
- Robin Delabays, Giulia De Pasquale, Florian Dörfler, and Yuanzhao Zhang. Hypergraph reconstruction from dynamics. *Nature Communications*, 2025.
- Urban Fasel, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A*, 478(2260):20210904, 2022.
- Kai Fukami, Takaaki Murata, Kai Zhang, and Koji Fukagata. Sparse identification of nonlinear dynamics with low-dimensionalized flow representations. *Journal of Fluid Mechanics*, 2021.
- Lloyd Fung, Urban Fasel, and Matthew Juniper. Rapid bayesian identification of sparse nonlinear dynamics from scarce and noisy data. In *Proceedings A*. The Royal Society, 2025.

- Daniel R Gurevich, Patrick AK Reinbold, and Roman O Grigoriev. Robust and optimal sparse regression for nonlinear pde models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(10), 2019.
- Jonathan Horrocks and Chris T Bauch. Algorithmic discovery of dynamic models from infectious disease data. *Scientific reports*, 2020.
- Kadierdan Kaheman, Steven L Brunton, and J Nathan Kutz. Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. *Machine Learning: Science and Technology*, 3(1):015031, 2022.
- Alan A. Kaptanoglu, Brian M. de Silva, Urban Fasel, Kadierdan Kaheman, Andy J. Goldschmidt, Jared Callahan, Charles B. Delahunt, Zachary G. Nicolaou, Kathleen Champion, Jean-Christophe Loiseau, J. Nathan Kutz, and Steven L. Brunton. Pysindy: A comprehensive python package for robust sparse system identification. *Journal of Open Source Software*, 7(69):3994, 2022. doi: 10.21105/joss.03994. URL <https://doi.org/10.21105/joss.03994>.
- Daniel A Messenger and David M Bortz. Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021.
- Daniel A Messenger, April Tran, Vanja Dukic, and David M Bortz. The weak form is stronger than you think. *arXiv preprint arXiv:2409.06751*, 2024.
- Robert K Niven, Laurent Cordier, Ali Mohammad-Djafari, Markus Abel, and Markus Quade. Dynamical system identification, model selection, and model uncertainty quantification by bayesian inference. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2024.
- Joshua S North, Christopher K Wikle, and Erin M Schliep. A bayesian approach for data-driven dynamic equation discovery. *Journal of Agricultural, Biological and Environmental Statistics*, 2022.
- Jeffrey Park, Craig R Lindberg, and Frank L Vernon III. Multitaper spectral analysis of high-frequency seismograms. *Journal of Geophysical Research: Solid Earth*, 92(B12):12675–12684, 1987.
- A Pearson and F Lee. On the identification of polynomial input-output differential systems. *IEEE Transactions on Automatic Control*, 30(8):778–782, 1985.
- AE Pearson. Aerodynamic parameter estimation via fourier modulating function techniques. Technical report, NASA, 1995.
- AE Pearson and JQ Pan. Frequency analysis via the method of moment functionals. Technical report, NASA, 1990.
- Patrick AK Reinbold, Daniel R Gurevich, and Roman O Grigoriev. Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Physical Review E*, 101(1):010203, 2020.
- Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science advances*, 3(4):e1602614, 2017.

- Benjamin P Russo and M Paul Laiu. Convergence of weak-sindy surrogate models. *SIAM Journal on Applied Dynamical Systems*, 23(2):1017–1051, 2024.
- Hayden Schaeffer and Scott G McCalla. Sparse model selection via integral terms. *Physical Review E*, 96(2):023302, 2017.
- Marvin Shinbrot. On the analysis of linear and nonlinear systems. *Transactions of the American Society of Mechanical Engineers*, 79(3):547–551, 1957.
- Robert Stephany and Christopher Earls. Weak-pde-learn: A weak form based approach to discovering pdes from noisy, limited data. *Journal of Computational Physics*, 506:112950, 2024.
- Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.
- David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 2005.
- Heinz Unbehauen and GP Rao. Continuous-time approaches to system identification—a survey. *Automatica*, 26(1):23–35, 1990.
- Marieke K van Vugt, Per B Sederberg, and Michael J Kahana. Comparison of spectral analysis methods for characterizing brain oscillations. *Journal of neuroscience methods*, 162(1-2):49–63, 2007.
- Sheng Zhang and Guang Lin. Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018.

Supplementary Information for “Fourier Weak SINDy: Spectral Test Function Selection for Robust Model Identification”

Zhiheng Chen

Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York, USA, 14850

ZC548@CORNELL.EDU

Urban Fasel

Department of Aeronautics, Imperial College London, London, United Kingdom

U.FASEL@IMPERIAL.AC.UK

Anastasia Bizyaeva

Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York, USA, 14850

ANASTASIAB@CORNELL.EDU

1. Multitaper Spectral Density Estimation

In this section, we review classical ideas from spectral analysis of signals (Stoica et al., 2005), which are used in the spectral density estimation component of our method. Given a right-sided random signal sequence $y(n)$ with length N , its power spectral density (PSD) $S(\omega)$ is defined as

$$S(\omega) = \lim_{N \rightarrow \infty} E \left\{ \frac{1}{N} \left| \sum_{n=0}^{N-1} y(n) e^{-i\omega n} \right|^2 \right\} \quad (1)$$

where E denotes expectation and \mathbf{i} is the imaginary unit. Note that $\sum_{n=0}^{N-1} y(n) e^{-i\omega n}$ is the discrete-time Fourier transform of the right-sided $y(n)$; therefore, $S(\omega)$ is a measurement of the amount of energy in $y(n)$ at frequency ω , and the function S represents the distribution of $y(n)$'s energy spanning a continuous spectrum of frequencies, which is a direct result from Parseval's theorem.

Direct computation of the PSD is not possible since it requires an infinite number of ensembles for $y(n)$, and thus spectral density estimations (SDEs) are needed. A common approach to SDE is via the periodogram, defined as

$$\hat{S}_P(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} y(n) e^{-i\omega n} \right|^2. \quad (2)$$

The periodogram by itself cannot provide decent spectral density estimations due to its high variance. To obtain a consistent PSD estimate, the multitaper method averages modified periodograms obtained using a set of orthogonal window functions (tapers); a typical set of tapers is the Discrete Prolate Spheroidal (Slepian) Sequences (Thomson, 2005; Slepian, 1978). Denote the j -th Slepian sequence as g_j , and denote the bandwidth normalized by the sampling rate as W , then g_j can be defined by its convolution with the sinc function:

$$\sum_{m=0}^{N-1} \frac{\sin(2\pi W(n-m))}{\pi(n-m)} g_j(m) = \lambda_k g_j(n), \quad n, j = 0, 1, 2, \dots, N-1, \quad (3)$$

and g_j 's can be obtained by solving the eigenvalue problem for the sinc function matrix. Then the j -th modified periodogram is constructed by first tapering the signal with g_j , then computing the squared magnitude of its discrete-time Fourier transform:

$$\hat{S}_j(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} g_j(n) y(n) e^{-i\omega n} \right|^2 \quad (4)$$

There are approximately $2NW$ sequences with eigenvalues close to unity, and thus we choose $M = 2NW$ tapers. Then the PSD estimate from the multitaper method is the average of the M modified periodograms:

$$\hat{S}_{MT}(\omega) = \frac{1}{M} \sum_{j=0}^{M-1} \hat{S}_j(\omega) \quad (5)$$

Note that the sinc function is the impulse response of an ideal lowpass filter with passband $[-W, W]$, and thus choosing a larger bandwidth will result in the tapers having a wider frequency content, causing the tapers to have a stronger frequency-averaging effect that can be visualized if we convert the Fourier transform of the time-domain multiplication in (4) into frequency-domain convolution. Therefore, if we choose a smaller and more concentrated bandwidth, then the PSD estimate will have more spectral leakage but higher resolution; conversely, if we choose a larger and more distributed bandwidth, then the PSD estimate will have less leakage but lower resolution.

2. Benchmark Systems

In this section we define the structure, parameters, and initial conditions for ODE systems used to evaluate our equation learning method, with equations for the hyperchaotic benchmark systems chosen from Gilpin (2023). The Lorenz system has governing differential equations

$$\dot{x} = \sigma(y - x) \quad (6a)$$

$$\dot{y} = x(\rho - z) - y \quad (6b)$$

$$\dot{z} = xy - \beta z \quad (6c)$$

where we use $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, and initial condition $[20 \ 12 \ -30]^T$ for benchmarking.

The Lotka-Volterra equations are

$$\dot{x} = 3x - \beta xy \quad (7a)$$

$$\dot{y} = -6y + \beta xy \quad (7b)$$

where we use $\beta = 1$ and initial conditions $[1 \ 1]^T$.

The hyperchaotic Lorenz system has governing differential equations

$$\dot{x} = a(y - x) + w \quad (8a)$$

$$\dot{y} = -xz + cx - y \quad (8b)$$

$$\dot{z} = -bz + xy \quad (8c)$$

$$\dot{w} = dw - xz \quad (8d)$$

where we use $a = 10$, $b = 2.667$, $c = 28$, $d = 1.1$, and initial conditions $[5 \ 8 \ 12 \ 21]^T$.

The hyperchaotic Jha system has governing differential equations

$$\dot{x} = a(y - x) + w \quad (9a)$$

$$\dot{y} = -xz + bx - y \quad (9b)$$

$$\dot{z} = xy - cz \quad (9c)$$

$$\dot{w} = -xz + dw \quad (9d)$$

where we use $a = 10$, $b = 28$, $c = 8/3$, $d = 1.3$, and initial conditions $[0.1 \ 0.1 \ 0.1 \ 0.1]^T$.

3. Simulation Results for Higher Order Polynomial Dictionaries

In addition to numerical experiments presented in the main text, we conducted numerical experiments with higher-order polynomial dictionaries, shown in Fig. 1.

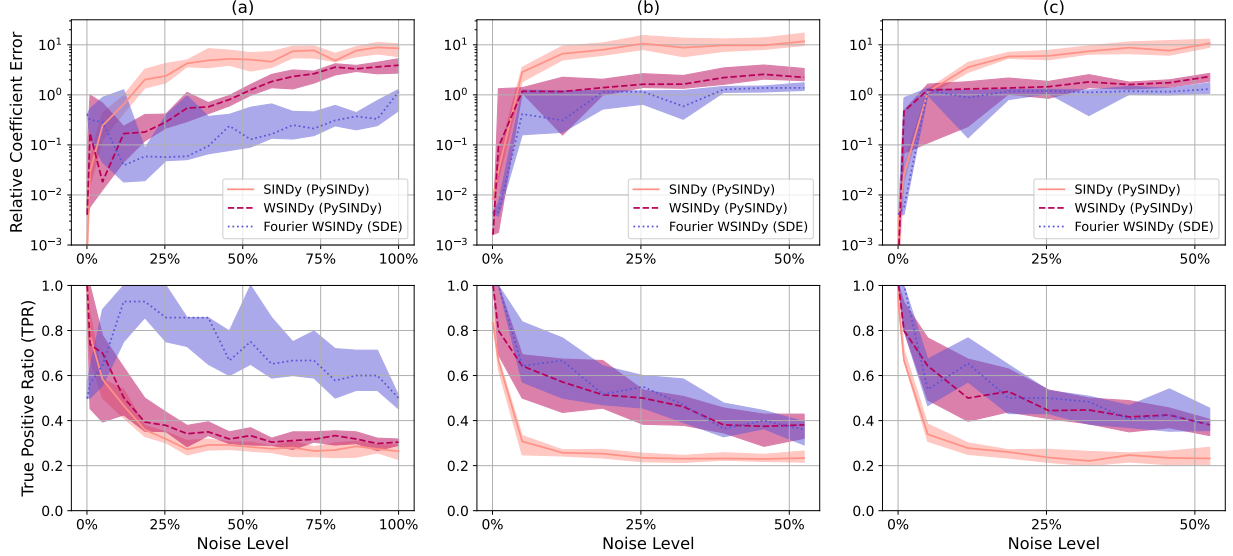


Figure 1: Equation learning results comparisons of (a) Lorenz system with fifth-order polynomial library, (b) hyperchaotic Lorenz system with third-order polynomial library, and (c) hyperchaotic Jha system with third-order polynomial library.

Note that in the experiments on the bump-function weak SINDy for the hyperchaotic Jha and hyperchaotic Lorenz systems with the third-order polynomial library, we choose sequentially-thresholded ridge regression (STRR) proposed by Rudy et al. (2017) rather than sparse relaxed regularized regression as STRR offers better accuracy in this experimental setting.

4. Simulation Results for Trajectory Error

We conducted numerical simulations that compare the trajectory errors of regular SINDy, bump-function weak SINDy, and Fourier weak SINDy. The relative trajectory error is defined as

$$\varepsilon_2 = \frac{\|\hat{X} - X_{\text{true}}\|_F}{\|X_{\text{true}}\|_F} \quad (10)$$

where \hat{X} is the simulated trajectory of a learned model and X_{true} is the simulated trajectory of the true model. All tests are performed on the Lorenz system with second-order polynomial library. Out of 200 equation learning results from each SINDy variant (up to 30% noise level), the regular SINDy and bump-function weak SINDy both learned 12 unstable dynamics models, while all dynamics models learned by Fourier weak SINDy are stable. The simulation results are shown in Fig. 2.

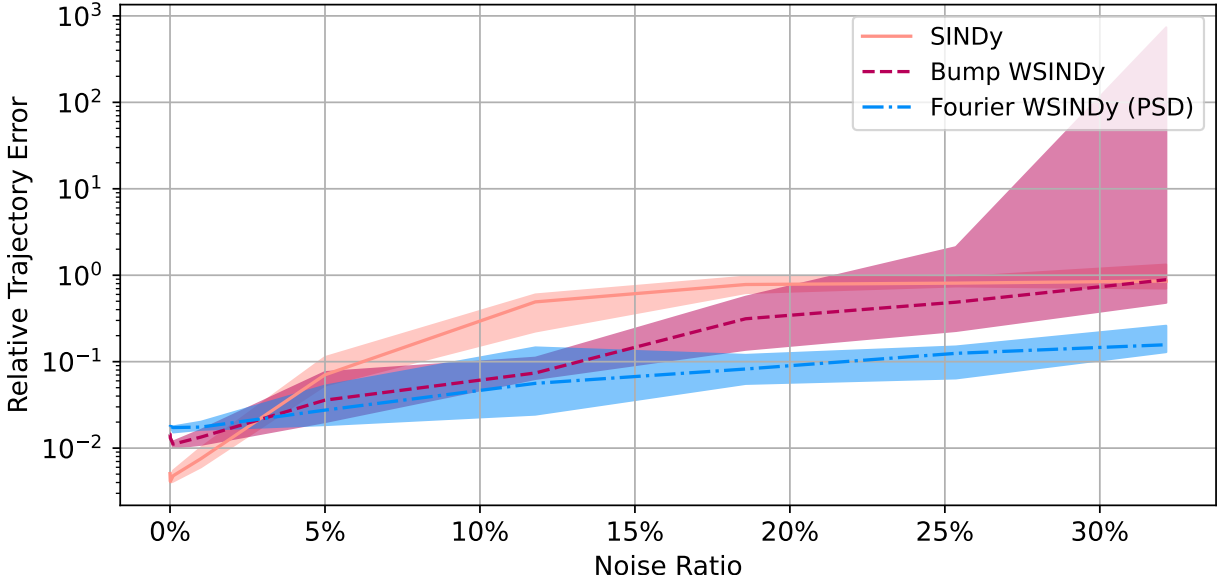


Figure 2: Trajectory error comparisons for regular SINDy, bump-function weak SINDy, and Fourier weak SINDy.

References

- William Gilpin. Chaos as an interpretable benchmark for forecasting and data-driven modelling, 2023. URL <https://arxiv.org/abs/2110.05266>.
- Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science advances*, 3(4):e1602614, 2017.
- David Slepian. Prolate spheroidal wave functions, fourier analysis, and uncertainty—v: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.
- Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.
- David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9): 1055–1096, 2005.